2018

Using atmospheric O2 and CO2 measurements to determine the stoichiometry of photosynthesis and respiration in a temperate forest

Margaret Marie Conley
Bowdoin College, mconley2712@gmail.com

Using atmospheric O_2 and CO_2 measurements to determine the stoichiometry of photosynthesis and respiration in a temperate forest

An Honors Paper for the Department of Physics and Astronomy

By Margaret Marie Conley

Bowdoin College, 2018

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Summary

As humans continue to emit CO_2 into the atmosphere via the combustion of fossil fuels, an understanding of carbon sources and sinks is critical to projections of future atmospheric CO_2 levels and climate change. CO_2 emitted into the atmosphere has three potential destinations. It may be sequestered in the terrestrial biosphere via photosynthesis, it may dissolve in the ocean, or it may remain in the atmosphere, contributing to rising levels of atmospheric CO_2 [*Keeling et al.*, 1996]. By enhancing the natural greenhouse effect, rising CO_2 levels cause an increase in average surface temperatures, but the degree of warming depends on the abundance of atmospheric CO_2. Understanding how much of the CO_2 that humans emit will remain in the atmosphere and how much will be stored in biomass on land is therefore necessary to project future warming, ocean acidification, and other changes in the earth system.

Calculation of the land carbon sink requires an accurate value for the $O_2 : CO_2$ exchange ratio of photosynthesis and respiration, known as α_b [*Battle et al.*, 2000]. This stoichiometric ratio is the number of O_2 molecules produced (or consumed) per CO_2 molecule consumed (or produced). Because the ratio depends upon the type of organic matter be-

ing synthesized or oxidized, from simple glucose to more complex molecules, the average value of α_b for the terrestrial biosphere is not well constrained, with a range of potential values from 1.0 to 1.2 [*Keeling and Manning*, 2014]. Measurements of atmospheric O_2 and CO_2 in forested regions can help constrain the average value of the $O_2 : CO_2$ exchange ratio for the forest biome as a whole, since changes in O_2 and CO_2 in the air over time will be driven by either photosynthesis or respiration and will reflect a range of processes carried out by both plants and soil bacteria over the course of daily and seasonal cycles. However, because the atmosphere is constantly in motion, the ratio of $O_2 : CO_2$ in the air may also reflect the characteristics of the region from which the air moving through the forest originated, so time periods of analysis must be chosen to maximize local impact and minimize the effect of outside influence.

In this project, we investigate the stoichiometry of O_2 and CO_2 at Harvard Forest in Petersham, Massachusetts over a span of six years, considering the covariation of O_2 and CO_2 during 6-hour periods to determine an average value for the stoichiometric ratio of the activity of the terrestrial biosphere. This approach differs from traditional methods of measuring the $O_2 : CO_2$ exchange ratio using closed chambers or elemental analysis and provides a way to determine the value of this parameter averaged across seasonal cycles and species assemblages. Comparing measurements within and above the canopy and during nighttime and daytime periods, we examine the importance of changing atmospheric dynamics and canopy effects on the observed exchange ratio, and we look for potential seasonal cycles. After our analysis of the observed $O_2 : CO_2$ exchange ratio, we turn to the question of whether this value is truly representative of α_b. We use a simple model to determine the range of variability in CO_2 and O_2 mixing ratios expected from local influence alone and use this as a criteria to isolate periods dominated by local exchange. This analysis yields important insight into the average value and variability of α_b for temperate forests for use in calculation of the land carbon sink.

The remainder of this chapter provides background information on land carbon

sink calculations, measurement techniques for α_b, and atmospheric and seasonal dynamics. Chapter 2 details the process of data collection and calibration. Chapter 3 describes methods used to split the data into 6-hour periods for analysis, the sensitivity of the analysis to time period lengths and fitting processes, and the results for the average stoichiometric ratio. Chapter 4 details the use of spectral analysis to look for seasonal or other cycles in α_b, as well as a discussion of seasonal variability in biospheric and fossil fuel signals. In Chapter 5, we use a Lagrangian transport model to investigate changes in the source region of air arriving at Harvard Forest, and we develop a simple 1-D box model to discern the amount of CO_2 and O_2 variability attributable to local exchange alone. This analysis enables us to understand the relevance of the observed $O_2 : CO_2$ exchange ratio as a measure of α_b for the surrounding forest. Chapter 6 describes the results of filtering using friction velocity and CO_2 ranges to isolate periods of local influence, which yield a more accurate value of α_b than the entire dataset.

1.2 Background

1.2.1 Carbon sink calculations

Carbon sink calculations partition CO_2 uptake into land and ocean fluxes based on key differences between these processes. Terrestrial processes, including fossil fuel combustion, respiration, and photosynthesis, involve the coupled exchange of O_2 and CO_2. Oceanic uptake and release of O_2, however, is essentially decoupled from CO_2 fluxes, which are driven by the dissolution and chemical buffering of CO_2 in seawater [*Keeling and Manning*, 2014]. Measurements of atmospheric O_2 mixing ratios therefore help distinguish between the uptake of CO_2 by the terrestrial biosphere and the ocean. The land and ocean sinks can then be calculated using the known flux of CO_2 from cement production and fossil fuel combustion and the mass balance for atmospheric CO_2 and O_2 (measured relative to N_2) along with the stoichiometric ratios of biospheric processes

and fossil fuel combustion [*Battle et al.*, 2000]:

$$f_{\text{land}} = -\frac{\alpha_{\text{ff}}}{\alpha_{\text{b}}} f_{\text{fuel}} + \frac{1}{2.49}\frac{d(\text{O}_2/\text{N}_2)}{dt} \tag{1.1}$$

$$f_{\text{ocean}} = -\frac{d}{dt}\left(\frac{\text{CO}_2}{0.471} + \frac{\text{O}_2/\text{N}_2}{2.49}\right) - \frac{\alpha_{\text{b}} - \alpha_{\text{ff}}}{\alpha_{\text{b}}} f_{\text{fuel}} - f_{\text{cement}} \tag{1.2}$$

where f terms refer to CO_2 fluxes (negative for ocean and land denoting uptake, positive for cement and fossil fuels denoting emission), α_{ff} is the stoichiometric ratio of fossil fuel combustion, approximately 1.4, α_{b} is the $O_2 : CO_2$ exchange ratio for processes in the terrestrial biosphere, approximately 1.1, and the other coefficients are conversion terms. Because the abundance of O_2 in the atmosphere is much higher than that of CO_2, the fractional variations over time are much smaller, making geochemically informative measurements challenging [*Stephens et al.*, 2007]. In the past 20 years, developments in instrumentation have enabled precise measurement of O_2 over time in field conditions, allowing for an improved quantification of the land and ocean carbon sinks.

1.2.2 Measuring α_{b}

Calculation of the land carbon sink depends on both α_{b} and α_{ff} [*Battle et al.*, 2000]. The stoichiometric ratio for fossil fuel combustion is relatively well-defined, with a global average value of 1.4 [*Keeling*, 1988]. Although the ratio varies over time and space depending on the mix of fuels being burned, the change from year to year is minimal; from 1996 to 2008, the ratio ranged between 1.39 and 1.42 [*Steinbach et al.*, 2011]. The stoichiometry of photosynthesis and respiration for the terrestrial biosphere, however, is less certain. This ratio depends on the type of reaction taking place in plant cells, which ranges from synthesis of simple glucose during photosynthesis to production of more complex molecules such as lignin or nitrogen-containing molecules [*Seibt et al.*, 2004]. These processes have a range of stoichiometric ratios from 1.0 to 1.2 [*Keeling and Manning*, 2014]. The value of α_{b} varies between ecosystems and plant species as well

[*Gallagher et al.*, 2017], and contributions from soil respiration, which vary over time and between soil types [*Seibt et al.*, 2004; *Worrall et al.*, 2013], add to this variability. Further constraining the global average value of α_b will strengthen the confidence of our estimates for the land carbon sink.

Atmospheric measurements of O_2 and CO_2 may provide a useful method to determine α_b, because mixing ratios in the air of a forest will be driven by either photosynthesis or respiration and will reflect a range of processes carried out by both autotrophs and heterotrophs (both plants and soil bacteria) over the course of daily and seasonal cycles. Forest air mixing ratios therefore give an average value of α_b for the forest biome as a whole. Previous studies to determine α_b have used enclosed chambers to measure gas mixing ratios over time as well as analysis of the elemental composition of different types of organic matter. Original studies of soil carbon yielded a ratio of 1.1 ± 0.05 [*Severinghaus*, 1995]. A more recent analysis using the elemental composition of a global survey of soils and vegetation produced a slightly lower value of 1.04 ± 0.03 [*Worrall et al.*, 2013]. Measuring this ratio in the atmosphere itself may provide a better value of α_b for calculation of the land carbon sink by integrating the effects of seasonal changes and a heterogeneous landscape. An atmospheric measurement will also help validate the value currently in use for carbon sink calculations.

1.2.3 Atmospheric and seasonal dynamics

Atmospheric mixing ratios of O_2 and CO_2 depend on a combination of processes. The background abundances (the regional average mixing ratios of these gases in the atmosphere) are altered by exchange with the underlying landscape, including respiration, photosynthesis, and anthropogenic combustion of fossil fuels. Since these processes add and remove O_2 and CO_2 from the atmosphere according to their respective stoichiometric ratios, changes in the mixing ratios of these gases over time should reflect the stoichiometry of the dominant process driving O_2 and CO_2 variability. For a predominantly forested

area, this $O_2 : CO_2$ exchange ratio should therefore provide a measurement of α_b for the surrounding ecosystem.

Unfortunately, the ratio of $O_2 : CO_2$ in air does not exclusively reflect the local environment. The atmosphere is constantly in motion, even within the canopy of a forest, and the characteristics of a given parcel of air reflect its evolution as it crosses the landscape, exchanging with forests and cities along its path. This complicates the interpretation of the stoichiometric ratio of $O_2 : CO_2$ measured at a fixed point, because the region from which the air originated (the source region) may change over time. Variation in the mixing ratio of air due to changes in the source region (the background variability) is superimposed on variation due to local exchange, with the latter or both combined driving the changes in mixing ratio observed at a particular location. However, if for a certain time period the observed magnitude of CO_2 and O_2 variability does not exceed the range of variability potentially driven by local exchange alone, the ratio of $O_2 : CO_2$ will be an accurate measurement of α_b.

Both natural and anthropogenic processes that impact CO_2 and O_2 vary over time, leading to potential temporal variability in α_b. Photosynthesis and respiration vary daily and seasonally with forest growth cycles and the activity of soil bacteria. Fossil fuel combustion exhibits its own daily and seasonal cycles associated with rush hours, winter heating, and other energy usage patterns [*Steinbach et al.*, 2011]. Depending on the season and the amount of green space, urban fluxes of CO_2 can be similar to or exceed forest fluxes and can therefore have a wide influence on surrounding regions [*Velasco and Roth*, 2010]. At Harvard Forest, the site studied in this project, the fossil fuel CO_2 signal is much smaller than local and regional forest signals during the summer, but the balance changes towards more fossil fuel influence during the winter [*Potosnak et al.*, 1999].

The dynamics of atmospheric exchange vary over time as well. Friction velocity, or u*, is a measure of turbulence that describes the vertical flux of horizontal momentum [*Stull*, 2006]. During winter, when the canopy is leafless, friction velocities tend to be

higher, and this may influence the relative importance of local exchange and background variability. Surface exchange is limited to the atmospheric boundary layer, also known as the planetary boundary layer (PBL), the section of the atmosphere closest to the ground that interacts with the surface. At night, when the surface cools rapidly, a nocturnal inversion develops, creating a stagnant layer near the surface in which exchange is limited [*Stull*, 2006]. The forest canopy further impacts air exchange, creating a surface mixing layer in the area underneath the canopy [*Raupach et al.*, 1996]. These changes may lead to variability in the stoichiometry of CO_2 and O_2, so an investigation of changes in α_b over the course of a year will shed light on both the average value and the amount of deviation from that value due to seasonal changes.

Chapter 2

Data collection and methods

2.1 Experimental setup

The experimental setup is located in Harvard Forest, a 3700 acre Long Term Ecological Research site in the town of Petersham in north-central Massachusetts (42.5°N, 72°W). The site has rolling topography and a temperate climate, with a mixed hardwood and conifer forest dominated by red oak, red maple, black birch, white pine, and Eastern hemlock [*Harvard University*, 2011]. Harvard Forest has extensive research facilities, including five research towers. Our setup is mounted on the Environmental Measurement Station (EMS) eddy flux tower, a 30 m tower in the Prospect Hill Tract, north of Petersham and east of route 32 at 42.5377°N, 72.1714°W (Figure 2.1).

We measured atmospheric O_2 and CO_2 concentrations at the Harvard Forest EMS tower from September 2006 to January 2013. Sampling was interrupted in mid-2010 by hardware failures and resumed in May 2012. We sample air from two heights, with one intake at 7.5 m (within the canopy) and another at 29 m (about 5 m above the canopy). Air enters the system through downward-facing aspirated inlets, which helps reduce thermal fractionation [*Blaine et al.*, 2006] (Figure 2.2). The air is transported via tubing down the tower and into a nearby building which houses the analysis instruments (Figure 2.3).

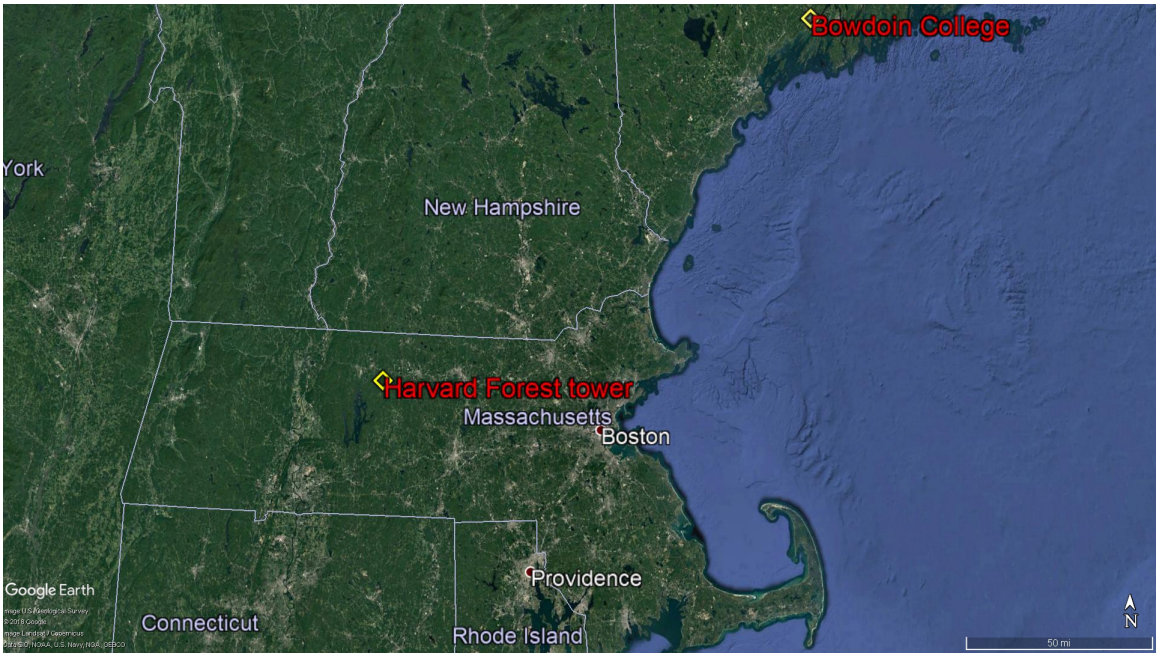

Figure 2.1: Map showing the location of the Harvard Forest EMS tower and Bowdoin College.

Figure 2.2: Photograph of the aspirated inlet at the high intake, designed to reduce thermal fractionation. Image credit: Mark Battle.

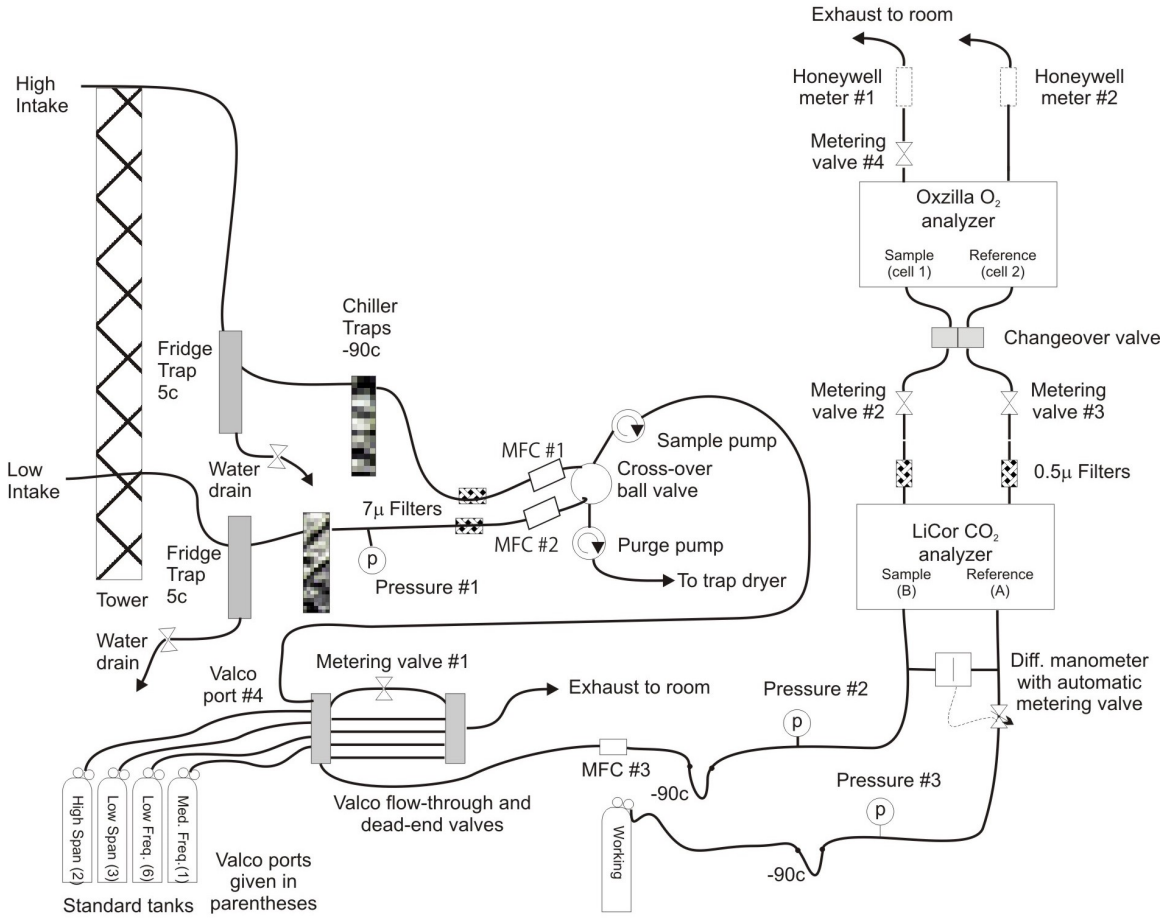

Figure 2.3: Diagram of the O_2 and CO_2 measurement system at Harvard Forest. Shows the path of air arriving from the tower as it passes through filters, chiller traps, and the two analyzers. Also shows the calibration tanks and the valve system used to switch between air sources. Credit: Mark Battle.

Following a design pioneered by *Stephens et al.* [2007], air flow is controlled using pumps and mass flow controllers. Once the air arrives in the building, it is cooled to remove water using a series of three coaxial stainless steel traps, with the first trap at 5°C and the next two traps at -90°C. As the air exits each trap, 7 μm filters remove particulates and ice crystals. The air is then routed to the O_2 and CO_2 analyzers. Only one airstream can be measured at a time, so the system alternates between the two heights, sending the airstream from the other height to the waste stream. This alternation occurs every 720 seconds, using a cross-over ball valve to switch the flow.

Measurements of O_2 are made with a Sable Systems Oxzilla II fuel cell analyzer, and CO_2 measurements with a LiCor LI-7000 NDIR analyzer. Each analyzer takes in two air streams, a sample and a reference. The mixing ratios in the sample air stream are measured relative to a reference "working" tank of dried air, which also passes through a chiller before entering the analyzer. Measurements are recorded every second. Because the concentration of the working tank may vary over time, the sample air is periodically diverted and replaced by air from one of four calibration tanks that have a known composition. These tanks allow us to correct the differential measurements of the sample stream for drift in either the instrument or the working tank composition. The air stream to be analyzed (standard tank or ambient air) is chosen using two 6-port selector valves. Sampling of ambient air is therefore interrupted for 18-minute calibration runs that occur about four times each day.

Before entering the LiCor CO_2 analyzer, the air stream passes through a mass flow controller and a differential manometer with an automatic metering valve. This ensures that the pressure within the cells of the LiCor remains balanced, because this can change the response of the analyzer. After passing through the LiCor, the air is directed through a 0.5 μm filter and a metering valve before entering the Oxzilla O_2 analyzer. The Oxzilla analyzer can also be affected by pressure differences as well as flow differences between the two cells, so there are metering valves both upstream and downstream of the sensor.

Because the cells may have different biases, we alternate the fuel cell into which each air stream flows using a changeover valve. The difference-of-differences is then calculated:

$$\begin{aligned} dd &= \frac{((Sa + \delta_1) - (Wt + \delta_2)) - ((Wt + \delta_1) - (Sa + \delta_2))}{2} \\ &= Sa - Wt \end{aligned} \tag{2.1}$$

where Sa is the sample mixing ratio, Wt is the working tank mixing ratio, and δ_1 and δ_2 represent the measurement errors associated with each fuel cell. This calculation accounts for differences in the response of the two cells. We use a procedure described by *Keeling et al.* [2004] to optimally set the frequency of changeovers. For our equipment, a switching period of 24 seconds maximizes measurement precision while minimizing the loss of measurements due to "dead time" (the time required for the instrument to stabilize after the pressure and flow disruptions associated with changeovers). The data collection set-up is remotely operated by a LabView® program, and data are transferred to a MySQL® database on a server at Bowdoin College.

2.2 Data calibration and processing

The raw data are processed to account for switches between air streams, to correct for instrumental drift, and to remove periods with data quality issues. Data collection alternates between atmospheric data and calibration runs, with the atmospheric analysis switching every 12 minutes between high and low intakes. Switching between atmospheric sampling and calibration tanks causes disruptions in pressure and flow. Thus, data collected in the first 432 seconds after a calibration run are discarded to allow the instrument to return to equilibrium. The same 432-second cut is applied following switches between the high and low intakes resulting in a data set of 288-second periods spaced 12 minutes apart, with periodic gaps from calibration runs.

O_2 data undergo a further processing step because of the changeovers between

the two cells. The 288 seconds of data from each sampling period are divided into twelve 24-second blocks corresponding to the changeovers between cells. The data for the last 10 seconds of each 24-second block (after the 14-second dead time) are averaged together, and the difference-of-differences is calculated for consecutive pairs of single-block averages, yielding 6 values for each 12-minute sampling period on the chosen intake. These 6 values are averaged to give a single O_2 value for that sampling period, with a time stamp corresponding to the middle of the time interval used in calculating the average.

CO_2 data are also processed by intake sampling period, but the process is simplified by the absence of changeovers. The 288 seconds of data are divided into the same twelve 24-second blocks as for the O_2 measurements. For consistency with the O_2 data, the same 10-second periods are used, but rather than calculating a difference-of-differences, consecutive groups of 10 data points are simply averaged together as a group of 20 points, yielding 6 values. These 6 values are again averaged together to yield a single data point with the same time stamp as the O_2 data point.

The calibration data from the standard tanks are processed differently from the atmospheric data. After 6 minutes of venting to flush stale gas from the lines, 6 minutes of measurements are recorded. Once the CO_2 value measured by the instrument has reached 70% of the change from atmospheric mixing ratios to standard mixing ratios, the remaining data are fitted with the function

$$[CO_2] = A + Be^{-C(t-D)} \tag{2.2}$$

This gives a value of CO_2 in the working tank relative to the standard tank. The O_2 measurements are processed using the same difference-of-differences treatment as the atmospheric analysis. After the CO_2 record shows that the atmospheric air has been displaced, the last three to five difference-of-differences values occurring after the transition are averaged.

Calibration runs generally occur every six hours and are used to track changes in the working tank and the instrumental response over time. The standard tanks have a wide range of O_2 and CO_2 values, allowing us to check the linearity of the LiCor and Oxzilla analyzers. One of the tanks is run only weekly to check for drifting of the other standard tanks. For the purpose of this research, the amount of drift in the standard tanks is small enough to have no detrimental effect on the results.

Poor quality data collected during periods of malfunctioning equipment or periods that lack calibration data are removed. Finally, the remaining data are adjusted using the calibration runs to correct for drift in the working tank and the instruments.

The calibration runs are also used to determine the instrumental error. For CO_2, the error is calculated from the scatter of data points about the fitted curve (Figure 2.4), whereas for O_2, the error is determined using the scatter in the three to five average difference-of-difference values recorded after the instrument reaches equilibrium with the standard tank. These errors varied over time (Figure 2.5). When fitting lines to plots of O_2 vs. CO_2, we split the data set into three time periods and used representative CO_2 and O_2 error values for each (Table 2.1).

Our final set of CO_2 and O_2 measurements was then augmented with friction velocity values measured at the Harvard Forest EMS tower [*Munger*, 2018]. The u* values were matched to individual CO_2 and O_2 values (where each value represents the result of the averaging process for each 12-minute sampling period). Friction velocity has the potential to influence the observed stoichiometry and is used in Section 6.2 to create subsets of data that reflect particular environmental parameters.

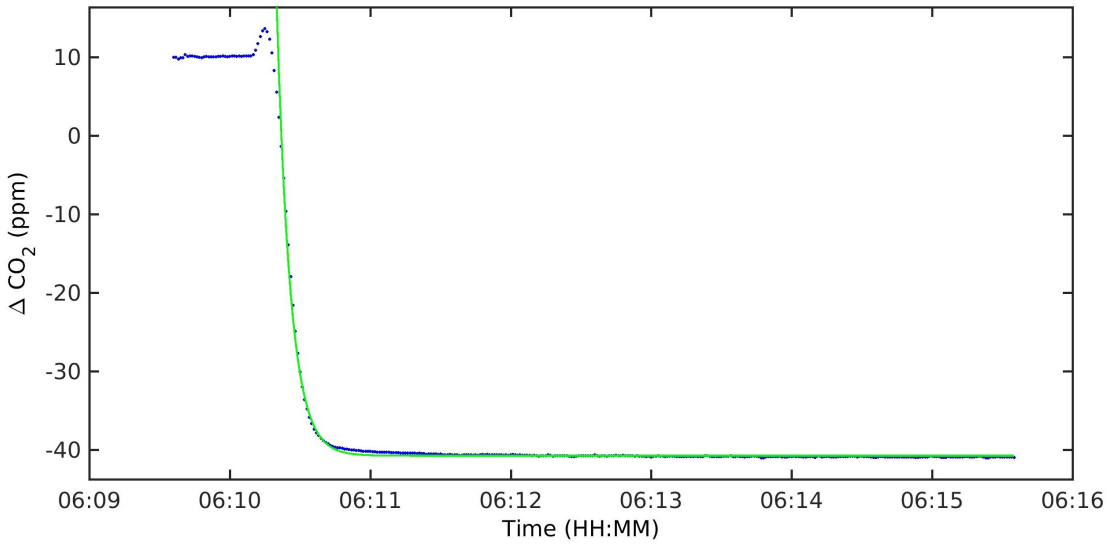

Figure 2.4: Plot of CO_2 vs. time for a calibration run, showing the curve (in green) fitted to the data points (in blue) about which the scatter is calculated to determine the instrumental error in CO_2 measurements.

Time period	Uncertainty in CO_2 (ppm)	Uncertainty in O_2 (μmol/mol)
Sept. 2006-July 2009	0.02	1.5
July 2009-July 2010	0.05	7.3
May 2012-Jan. 2013	0.13	11.7

Table 2.1: Uncertainty in CO_2 and O_2 measurements. Uncertainty differed between time periods because of variability in the stability of the instruments. The increase in uncertainty near the end of the study period may be due to problems with system gas tightness.

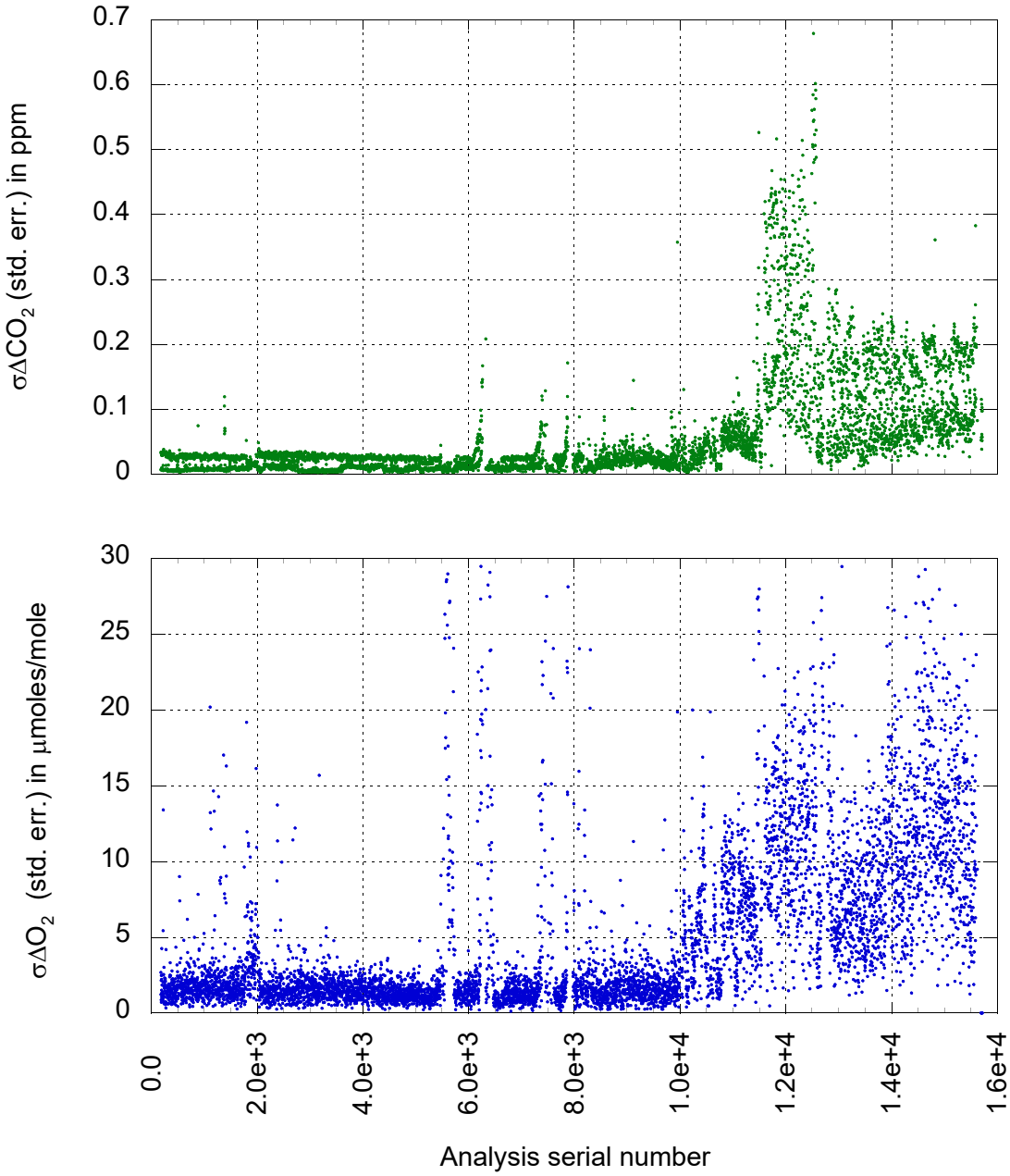

Figure 2.5: Error in CO_2 (top, in green) and O_2 (bottom, in blue) plotted against serial number, which corresponds to the chronological order of data points. Error varied over time and tended to increase significantly at the end of the study period. We speculate that this increase in error may be due to problems with system gas tightness.

Chapter 3

Analysis of the average stoichiometric ratio

3.1 Data categorization and analysis

We calculate α_{b} by analyzing the covariation of O_2 and CO_2 over time. For a plot of O_2 vs. CO_2, the slope of a linear fit gives the stoichiometric ratio of O_2 to CO_2, which for an atmospheric composition dominated by the terrestrial biosphere gives α_{b}. We aggregate the data into one-day periods and split them by intake height and time of day before calculating the slopes, which allows us to examine the effects of atmospheric and boundary layer processes on the observed stoichiometry. Each one-day period includes a daytime interval from 9am–3pm and a nighttime interval from 10pm–4am. The split between high and low intakes results in four subsets of data for each one-day period: day/high, day/low, night/high, and night/low. Examples of these plots are shown in Figure 3.1.

The duration of the aggregation period determines the effective "footprint" of the tower over that time period, where "footprint" refers to the area that contributes to the observed atmospheric composition at Harvard Forest. The size, composition, and

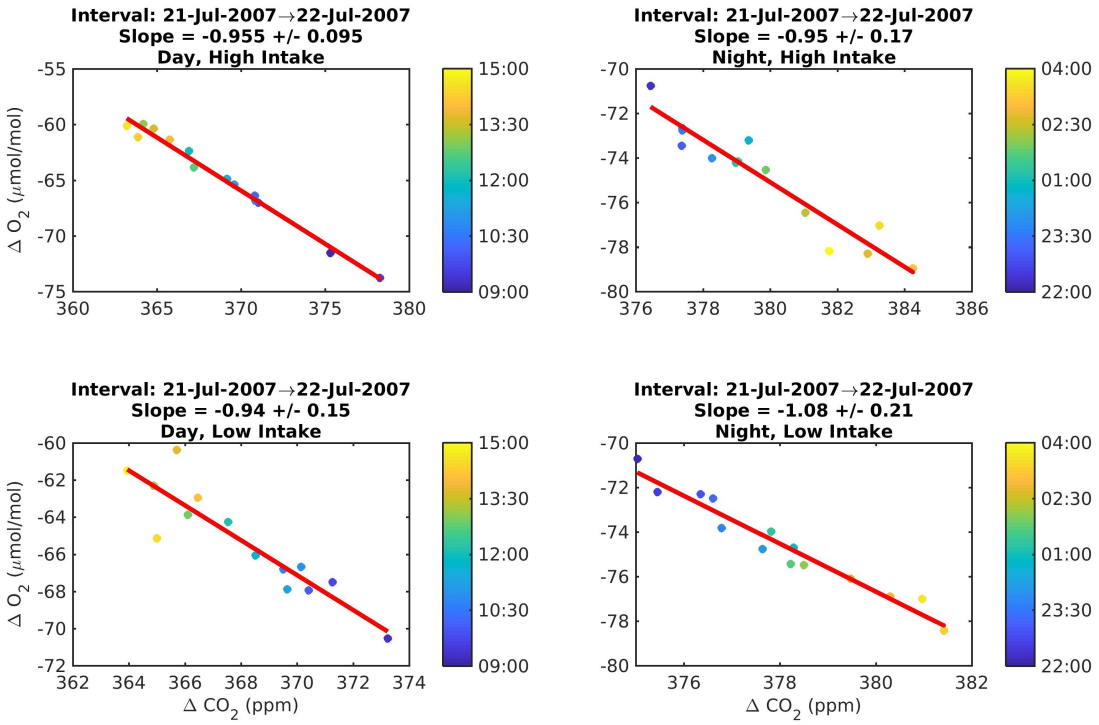

Figure 3.1: Example of scatterplots of O_2 vs. CO_2 for the four subsets of data: day and night at high and low intakes. These plots are for 9am–3pm on 21 July and 10pm–4am on 21–22 July in 2007. The color of the points indicates the time at which each data point was taken (in local time). Note the overall decrease in CO_2 and increase in O_2 during the daytime interval, driven by photosynthesis, and the overall increase in CO_2 and decrease in O_2 during the nighttime interval, driven by respiration. The red line shows the results of a Deming regression (described in Section 3.1), and the slope, corresponding to the $O_2 : CO_2$ exchange ratio, is indicated above each plot.

variability of this footprint are all important in determining the relevance of the observed $O_2 : CO_2$ ratio as a measurement of α_b. As parcels of air move across the landscape, passing over forests, cities, suburbs, and roads, they exchange O_2 and CO_2 with the surface. One-day periods (additionally separated into daytime and nighttime intervals) yield more informative slopes than longer spans of multiple days because the source region of incoming air is smaller for a shorter time period [*Gerbig et al.*, 2006]. The source region is also less variable over time and likely includes a landscape dominated by forest with a small influence from fossil fuels (see Section 5.2).

The daytime and nighttime intervals also affect the region of influence reflected in the observed slope value. We calculated the slope values for both 6-hour intervals (as described above) and 4-hour intervals (10am–2pm for day, 11pm–3am for night) to explore the impact of interval length on our results. Our definition of daytime and nighttime intervals is based on u*, the friction velocity. Nighttime is characterized by lower u* values due to higher stratification and limited vertical exchange, as air near the surface is trapped by a nocturnal stable boundary layer [e.g. *Stull*, 2006]. This may lead to a more local signal in the observed stoichiometry. In contrast, daytime is characterized by higher u* values indicating more exchange and vertical air movement, and consequently a potentially larger footprint influencing observed mixing ratios.

We used 24 years of u* data from the EMS tower at Harvard Forest to calculate an hourly climatology of average u* values for each hour of the day (Figure 3.2). The daily variation in u* shows a somewhat sinusoidal cycle, with higher u* values during the day, peaking around noon, and lower u* values throughout the night. We used this climatology to choose daytime and nighttime intervals that are dominated by the maximum and minimum u* values, respectively, avoiding transition periods because of the variability in sunrise and sunset times throughout the year. Although u* also shows pronounced seasonal cycles, with higher values in winter (Figure 3.3), as well as short-term variability due to high winds and storms, the daytime and nighttime splits are applied year-round

as a first-order approximation for stratified and well-mixed regimes. These regimes may in turn exhibit more or less local influence on the observed stoichiometry.

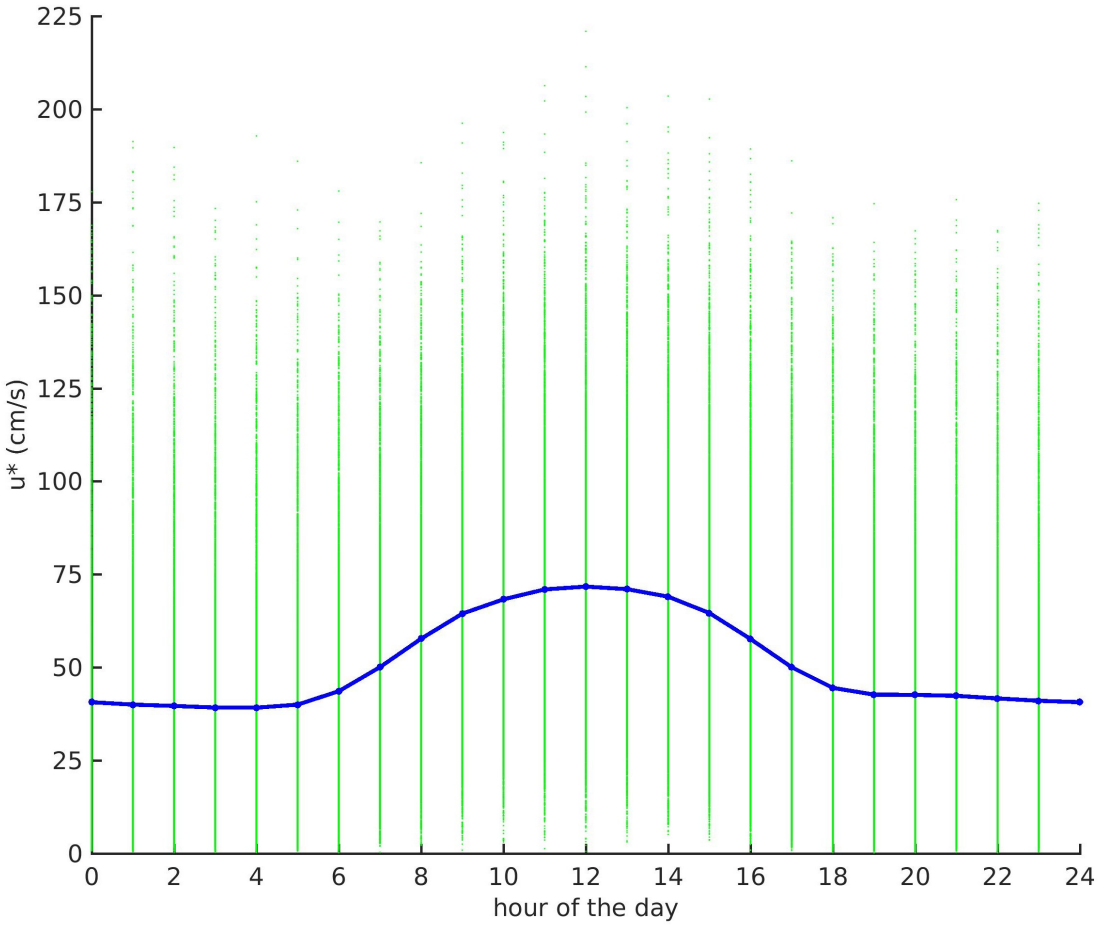

Figure 3.2: Hourly values of friction velocity (u*, cm/s). Green dots show hourly u* values for 1991-2015 measured at the Harvard Forest EMS tower, and the blue points and line show the hourly average. Although u* is highly variable, with a large amount of scatter for each hour of the day, standard errors on the hourly averages are too small to be seen on the plot, and the climatology is well constrained. Data courtesy of Harvard Forest online archives [*Munger*, 2018].

High and low intake heights may also exhibit different stoichiometry due to the "surface layer" created by the forest canopy. This surface layer, which exists within the larger planetary boundary layer (PBL), is characterized by a pronounced inflection point

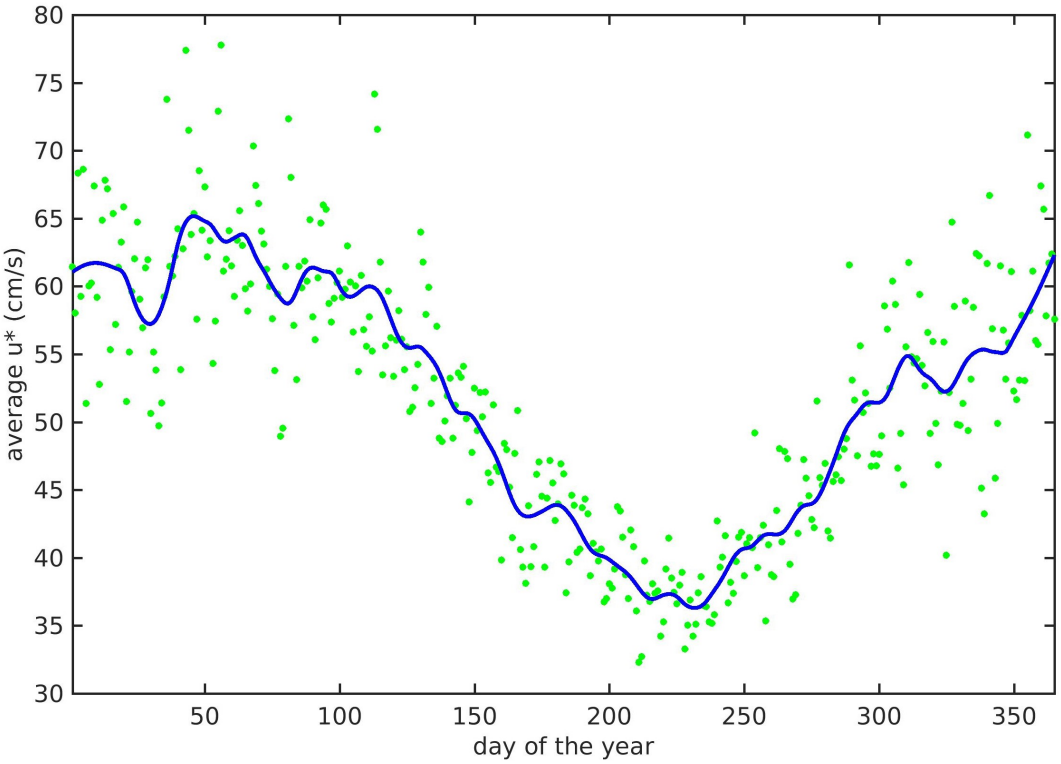

Figure 3.3: Seasonal variability in u* (cm/s). Green dots show the average u* value for each day of the year, calculated from hourly data for 1991-2015. The blue line shows the results of a loess smoothing function using a window of 36.5 days. Friction velocity tends to be higher and more variable in the winter months compared to the summer.

in velocity at the top of the canopy, accompanied by maximum shear and an increase in wind speed [*Raupach et al.*, 1996]. The high intake on the EMS tower, therefore, is in a transition zone between low velocities within the canopy and higher velocities above. Within the canopy, turbulence and coherent eddies transport and mix air parcels, contributing to vertical exchange [*Raupach et al.*, 1996]. At night, stable stratification can weaken mixing and lead to the accumulation of CO_2 and other gases within the canopy, especially at low points in the landscape [*de Araùjo et al.*, 2008]. Stratification may increase the relative influence of local signals at the low intake compared to the high intake. However, canopy waves can develop under stably stratified conditions, leading to vertical mixing, and rolling topography can further complicate the flow of wind within the canopy [*Lee*, 2000]. These dynamics may contribute to potential differences in stoichiometry between high and low intakes.

We use the slopes of linear fits to scatter plots of O_2 vs. CO_2 to calculate the $O_2 : CO_2$ exchange ratio. Slopes are calculated using a Deming regression, a type of orthogonal fit that takes into account significantly different errors in the x and y variables [*Deming*, 2011]. Before performing the regression, daytime or nighttime intervals with too few data points (fewer than three) are removed. Then an iterative process is used to remove outliers before calculating the final fit. After calculating an orthogonal fit to the data set along with the orthogonal residuals and the standard deviation of the residuals (σ_{res}), points with a residual greater than $4\sigma_{\text{res}}$ are cut from the data set, and the process is repeated until the data set converges and no points have a residual greater than $4\sigma_{\text{res}}$ (Figure 3.4). The slopes are negative because increases in CO_2 coincide with decreases in O_2, and vice versa.

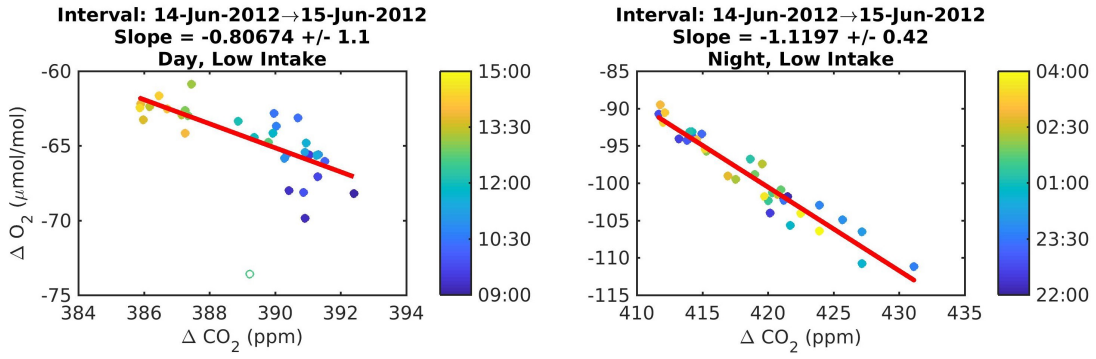

Figure 3.4: Scatterplots of O_2 vs. CO_2 for 14–15 June 2012 at the low intake. The empty circle denotes a point that was removed by the orthogonal residual cut and was not included in the final fit shown by the red line.

3.2 Averaging of slopes and sensitivity tests

Using the slopes for each of the 3760 6-hour periods, we calculate an average value for the $O_2 : CO_2$ exchange ratio separately for the low and high intakes during daytime and nighttime intervals, as well as an overall average including both intakes and intervals. To determine a representative average slope value, we examined the distribution of slopes. Based on histograms and normal probability plots that show the degree to which the distributions of slopes fit a normal distribution, the distributions for both the aggregation of all slopes and the four subsets have Gaussian cores but non-Gaussian tails (Figures 3.5 and 3.6). We expect the aggregation of all slopes in particular to be non-Gaussian, because it consists of four potentially different distributions of slopes for day/high, day/low, night/high, and night/low subsets.

Given the non-Gaussian tails, we employ an iterative averaging of the slope values. We average all slopes and calculate the standard deviation, then discard slopes more than $n\sigma$ away from the mean. We then recalculate the average and iterate until convergence, when no points remain outside the $n\sigma$ window. We considered both $n = 3$ and $n = 4$ to test the sensitivity of our results to this procedure. Results are summarized in Table 3.1.

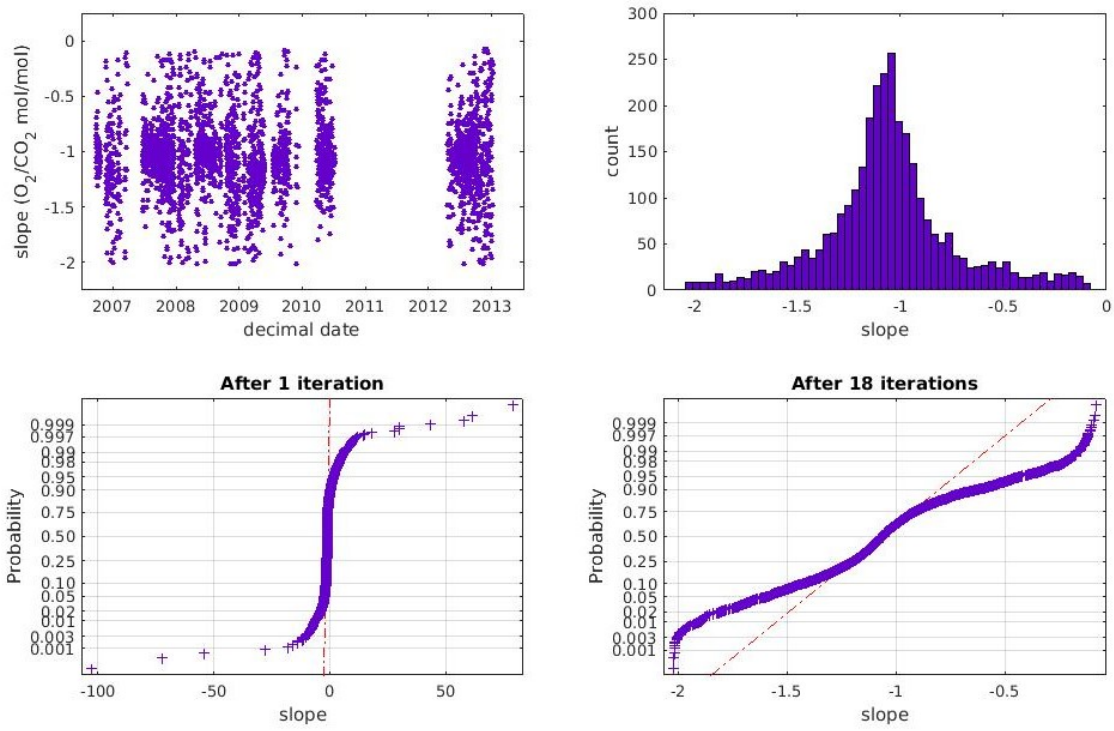

Figure 3.5: Distribution of slopes for the entire dataset after the iterative cut. The top left panel shows the iteratively cut slopes plotted against time in decimal years; gaps indicate periods where no data were collected. The top right panel shows a histogram of the iteratively cut slopes. The bottom left panel shows a normal probability plot for the entire dataset after the first iteration of the iterative cut, whereas the bottom right shows the same plot after the iterative cut was completed, demonstrating the reduction of the dataset's non-Gaussian tails. The dotted lines in the normal probability plots indicate the point pattern expected for a normal distribution. Summary statistics are in Table 3.1.

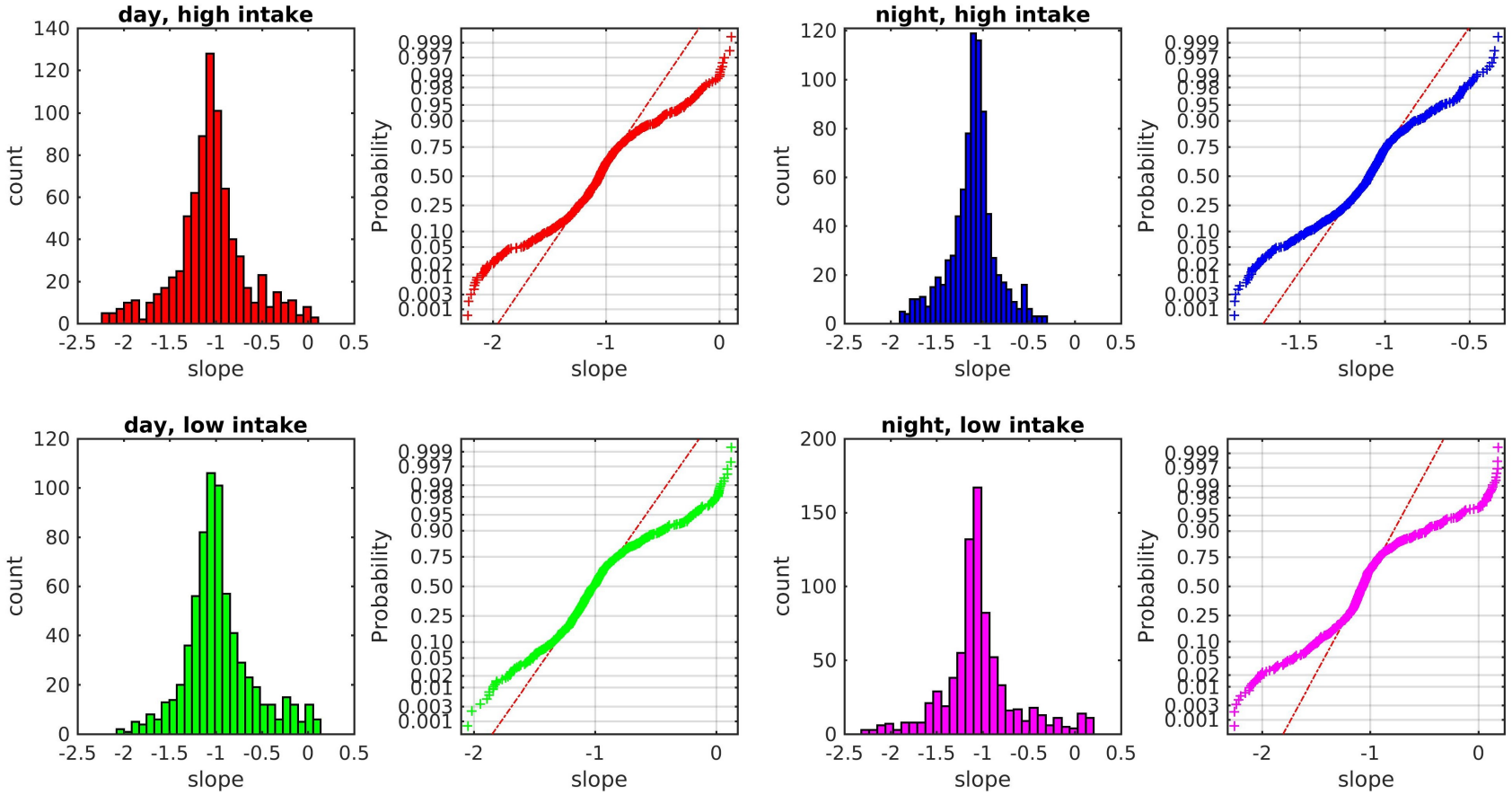

Figure 3.6: Histograms and normal probability plots for the four data subsets. Each normal probability plot corresponds to the histogram to its left. Although the iterative cut partially normalizes the distributions, there are still non-Gaussian tails, visible in the normal probability plots, in which the dotted line indicates the point pattern expected for a normal distribution. For this reason, we include both the means and the medians for comparison in the summary statistics in Table 3.1.

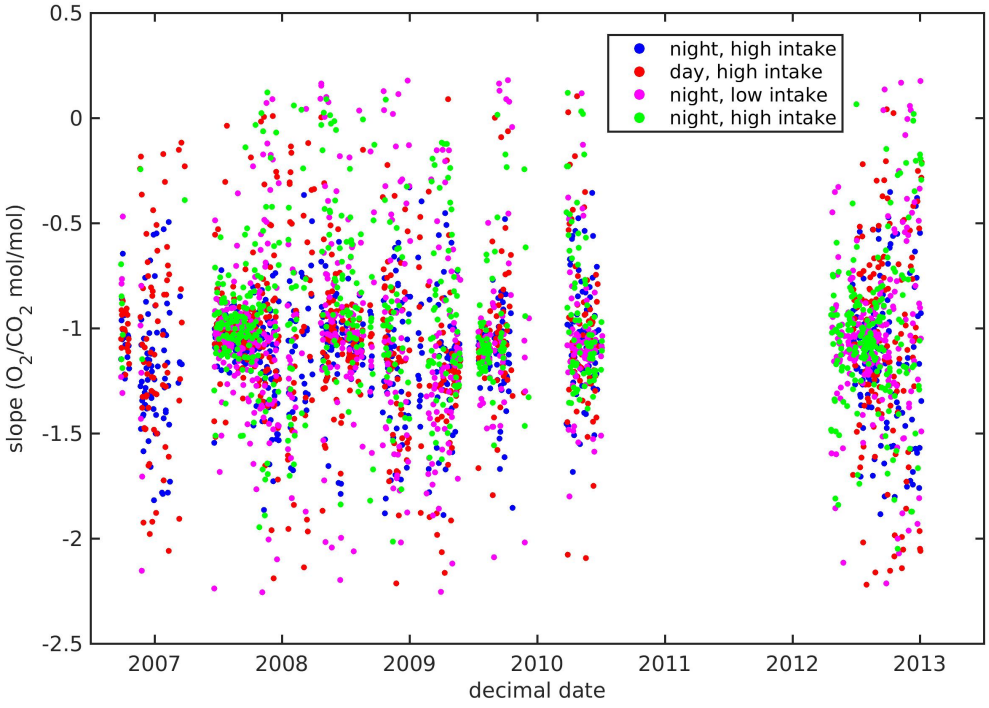

Figure 3.7: Plot of slopes vs. time in decimal date. Point colors indicate the time interval and intake height for each data point. The increase in spread towards the end of the study period may be a reflection of the increase in measurement error shown in Figure 2.5.

After a 3σ cut, the difference between the mean and median values is smaller, indicating that the cut reduces the non-Gaussian tails of the data set, although the mean and median do differ, especially for some subsets of the data such as day/low. For a 4σ cut, the difference between mean and median values is larger, and although the median values are similar to those produced by the 3σ cut, the mean values are quite different. This justifies the use of a relatively narrow 3σ window to remove outliers and normalize the distribution of slopes.

Although we considered other techniques to remove outlier slope values, none proved able to systematically remove poor fits. The error in the slope calculated during the regression process was highly variable, with some clearly flawed fits achieving low uncertainty due to an anomalously large range in O_2 or CO_2 (Figure 3.8). In other cases, an isolated point with an anomalously large O_2 or CO_2 value had a disproportionate impact on the slope (Figure 3.9). The cause of these high O_2 and CO_2 values is unknown. Values of chi-squared for the slopes proved similarly variable, with no clear trend of higher values for poor fits. The large number of slope plots (3760 total) meant that we were unable to identify obviously inappropriate fits by visual inspection.

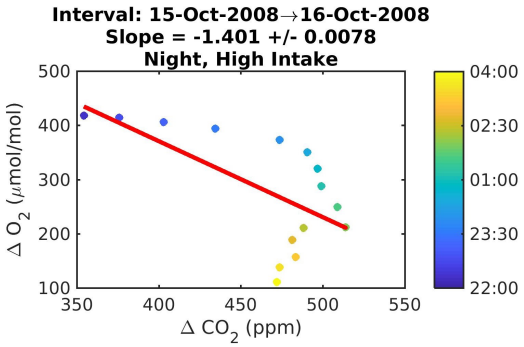

Figure 3.8: Scatterplot of O_2 vs. CO_2 for the nighttime interval on 15–16 October 2008. Although the linear fit is obviously inappropriate for this plot, the uncertainty in the regression is low. This is due to the anomalously large ranges in CO_2 and O_2, which are much higher than the ranges characteristic of the majority of the plots.

Interval length	Iterative cut	Data set	mean	median	SD	SEM	N	N cut	N iter
6-hour	3σ	All data	-1.06	-1.07	0.33	0.01	3058	702	18
		Day, high intake	-1.06	-1.06	0.39	0.01	804	149	15
		Day, low intake	-0.97	-1.02	0.37	0.01	703	194	20
		Night, high intake	-1.11	-1.09	0.26	0.01	819	165	14
		Night, low intake	-1.04	-1.07	0.41	0.01	798	128	10
4-hour	3σ	All data	-1.03	-1.06	0.45	0.01	3009	632	23
		Day, high intake	-0.99	-1.05	0.54	0.02	781	138	11
		Day, low intake	-0.92	-0.99	0.55	0.02	692	168	12
		Night, high intake	-1.10	-1.09	0.36	0.01	806	164	22
		Night, low intake	-1.08	-1.08	0.38	0.01	723	169	19
6-hour	4σ	All data	-0.98	-1.06	0.60	0.01	3488	271	13
		Day, high intake	-0.99	-1.05	0.76	0.03	904	49	7
		Day, low intake	-0.80	-0.99	0.88	0.03	827	70	15
		Night, high intake	-1.08	-1.09	0.49	0.02	913	71	12
		Night, low intake	-0.99	-1.06	0.59	0.02	851	74	16

Table 3.1: Table of average values for all slopes and the four data subsets, including mean, median, standard deviation (SD), standard error of the mean (SEM, defined as $\mathrm{SD}/\sqrt{\mathrm{N}}$), number of slopes values (N), number of slope values cut during the iterative process (N cut), and the number of iterations before convergence (N iter). Shown are the statistics for two interval lengths, 6 hours and 4 hours, as well as two iterative cutting windows, 3σ and 4σ, to test the sensitivity of the resulting average values to those parameters. Because the 4σ window results in a larger discrepancy between mean and median values, and the 4-hour interval increases uncertainty without significantly changing the average slope values, we concluded that the 6-hour, 3σ combination provides the best results.

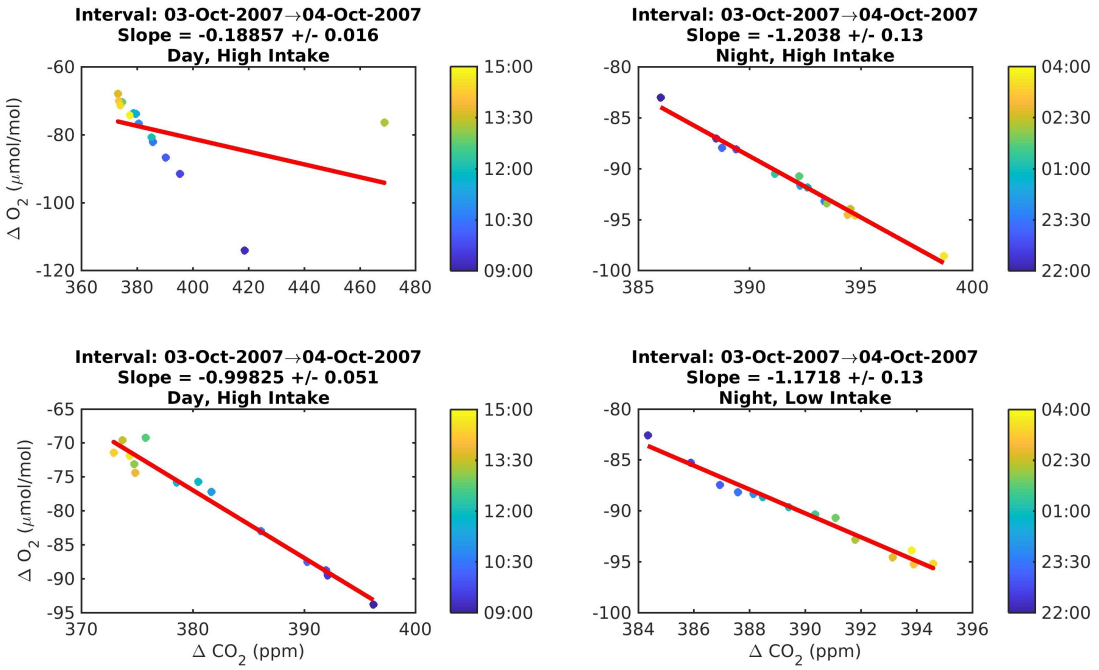

Figure 3.9: Scatterplots of O_2 vs. CO_2 for 3–4 October 2007. In the "Day, High Intake" plot, an anomalously high CO_2 value resulted in a non-representative fit without much increasing the uncertainty, while the other subsets, whose linear fits are much more representative of the trend in the data, have higher uncertainties. Similar to Figure 3.8, a point with a much higher CO_2 or O_2 value than the rest of the data causes problems for the linear regression and results in low uncertainties that do not reflect the actual quality of the fit.

We also compared statistics for 6-hour and 4-hour daytime and nighttime intervals, with the nighttime interval centered about 1am and the daytime interval centered about noon. Besides testing the robustness of the average slope value, this comparison helps us determine whether a shorter time period might significantly enhance the local signal in the slope values by further reducing variability in the source region. The 4-hour interval data set has fewer slope values overall, because the shortened interval reduces the number of available data points. This reduction of data points resulted in slightly larger standard deviations for all subsets of the data except night/low. For all subsets, the mean values for the 4- and 6-hour intervals are within one standard deviation of each other, and the median values differ even less. Because there is no notable change in the

average slope values, we concluded that 4-hour intervals simply reduce the number of data points and the statistical power of the analysis without altering the value of the $O_2 : CO_2$ exchange ratio. Therefore, we use 6-hour intervals for our calculation of α_b. Histograms, normal probability plots, and scatterplots for all slopes and the four subsets are shown in Figures 3.5, 3.6, and 3.7.

3.3 Discussion of results for the average stoichiometric ratio

Based on our analysis of the distribution of slope values, we believe that the mean provides the value of the stoichiometric ratio most representative of the underlying ecosystem processes. The iterative cutting process at least partly normalizes the slope distributions, and the medians and means are consistent within less than one standard deviation. In addition, the relative magnitudes of the means for the four data subsets (low vs. high and day vs. night) are consistent both for changes in the daytime and nighttime interval length and changes in the iterative fitting window. Mean and median slope values for all slopes and the four subsets are shown in Table 3.1.

Our all-data slope value of -1.06 $\pm$ 0.01 (taking the absolute value of the slope to obtain the exchange ratio) is in agreement with the values for α_b calculated by both *Severinghaus* [1995] and *Worrall et al.* [2013], although it is in the low end of the range reported by Severinghaus. This value is also lower than the standard value of 1.1 used in most calculations of the land carbon sink [*Keeling and Manning*, 2014]. If our measurements reflect mainly local influences and are representative of a predominantly forested landscape, they support the use of a slightly lower value of α_b (1.05–1.07) in such calculations. Even if our data represent a mix of forest fluxes with a small amount of fossil fuel influence, the addition of fossil fuels would raise the value of α_b above its true value, in the direction of α_{ff} (1.4). This gives further support to an adjustment in the value of

α_b towards 1.05.

The slope values for the four subsets of data provide a perspective on the influence of atmospheric processes in determining observed slope values. The night/high subset has the steepest slope, with a mean of -1.11 ± 0.01, followed by day/high (-1.06 ± 0.01), night/low (-1.04 ± 0.01), and day/low (-0.97 ± 0.01). The difference between day and night slopes, with more negative values at night for the same intake height, could result from several factors. Fossil fuel influence may be more apparent at night, when gas fluxes from forests are reduced but fossil fuel combustion decreases only slightly from daytime values (see Figure 5.7). Steeper nighttime slopes may also reflect changes in the dominant plant physiological processes occurring in the forest, such as the production or respiration of more nitrogen-rich molecules with higher stoichiometric ratios [*Keeling and Manning*, 2014]. Soil respiration may influence the slopes observed at night, especially at the low intake. Although stably stratified conditions at night may enhance local influence, periodic gusts of wind may nullify this effect. Shallower slope values during the day may, in turn, reflect an increased influence of forest fluxes relative to fossil fuel influence, pulling the exchange ratio closer to 1.05, or the dominance of photosynthesis over respiration altering the exchange ratio.

The difference between high and low intakes likely results from the impact of the canopy surface layer. Reduced air velocities within the canopy may enhance local influence and drive slope values towards α_b. At night, stratification may increase the isolation of the high intake from the low intake and result in background variability rather than local influence dominating the signal at the high intake. The day/low slope is particularly shallow compared to the other subsets and the average of all slopes, at the low end of values for the $O_2 : CO_2$ exchange ratio calculated by *Worrall et al.* [2013]. In contrast to the atmospheric boundary layer, which becomes stably stratified near the surface at night, the forest canopy can be more stable during the day than at night, because it is warmed from above by incoming sunlight, and this may drive the exchange

ratio towards smaller values for the day/low subset. At night, with cooler air above the treetops, upside-down thermals can mix the air within the canopy [e.g. *Stull*, 2006; *Jacobs et al.*, 1992], perhaps leading to relatively steeper slope values for the night/low subset.

Chapter 4

Seasonal cycles in atmospheric stoichiometry

4.1 Spectral analysis of slopes

We used spectral analysis to look for the presence of cycles in the slope values. Because our data are not evenly spaced, we used Lomb-Scargle periodograms to quantify the temporal structure of the data [*Press et al.*, 2007]. We created a power spectrum for all 6-hour slopes together, as well as spectra for the four subsets of data separately. Slopes were assigned times based on the approximate midpoint of each 6-hour period. The resulting periodograms include a power spectrum and an estimate of the significance of any peaks.

The four subsets and the combined slopes exhibit several different periodicities (Figures 4.1 and 4.2). Although the power of the peaks is relatively low overall, several peaks are significant at the $\alpha = 0.05$ level. For all slopes combined, there is a broad peak encompassing frequencies from 0.4 to 1.0 cycles per year (Figure 4.3) with one significant spike, indicating a one- to two-year cycle. For the night/low and day/high subsets, a significant peak appears at approximately 18 cycles/year, corresponding to a period of

about three weeks. This peak is significant at the $\alpha = 0.01$ level for the day/high slopes. The night/high subset has a significant peak at 1 cycle/year, while the day/low subset exhibits a peak at 0.6 cycles/year; cycles with frequencies of 18 are not apparent in these two data subsets, nor for all slopes combined.

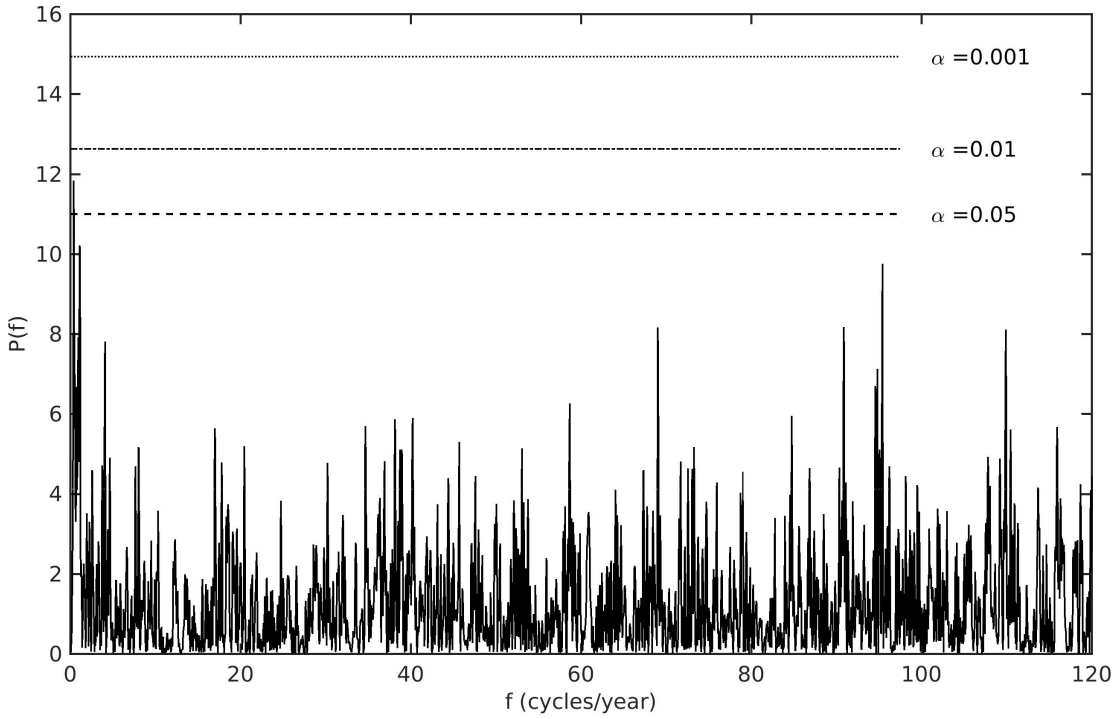

Figure 4.1: Lomb-Scargle periodogram for the entire set of slopes (after the iterative cut described in Section 3.2), showing the power spectrum for a range of frequencies in cycles/year. Dashed lines indicate the power thresholds for the indicated significance levels. For all slopes combined, there is evidence of a low frequency peak with no significant peaks at higher frequencies.

Several of these periodicities broadly correspond to a yearly seasonal cycle. This cycle is clearly apparent in the night/high subset, and the all slopes and the day/low subset have periodicities close to one cycle/year. A comparison of the average slope value for winter (slopes from the months of December, January, and February) and summer (from the months of June, July, and August) reveals significantly different average values

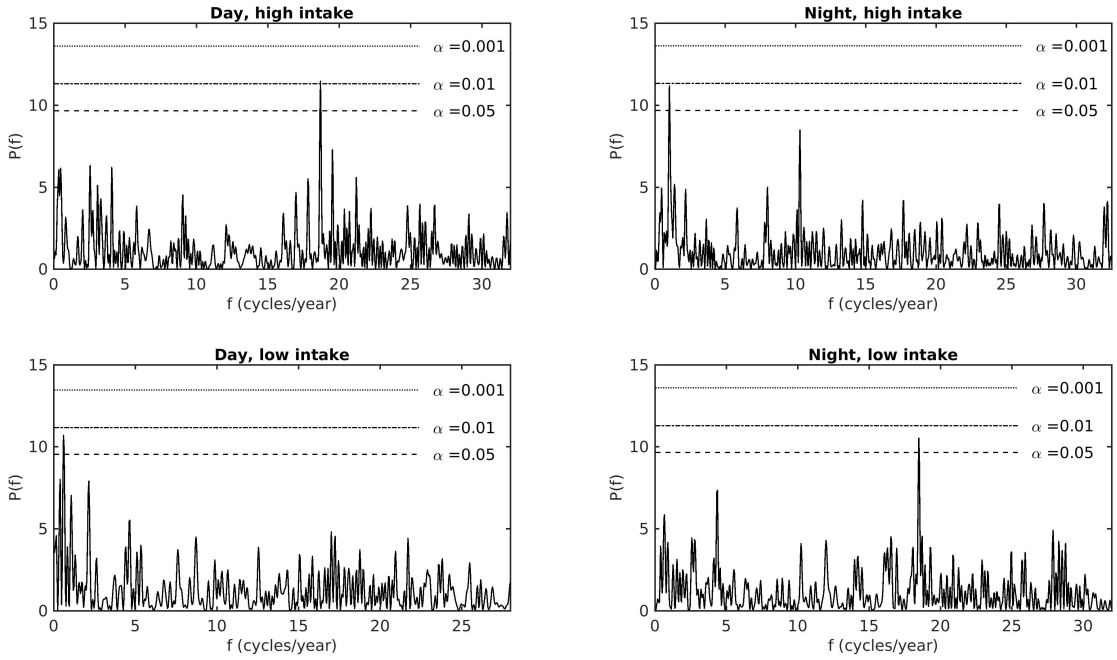

Figure 4.2: Lomb-Scargle periodograms for the four subsets of slopes (after the iterative cut), showing power spectra for a range of frequencies in cycles per year. Dashed lines indicate the power thresholds for the indicated significance levels. For day/high and night/low subsets, significant peaks occur at a frequency of approximately 18 cycles/year (a period of three weeks), which we speculate may be due to the timing of calibration runs rather than an actual environmental signal. The day/low and night/high subsets show significant peaks at 0.6 and 1.0 cycles/year, close to a seasonal period of one year.

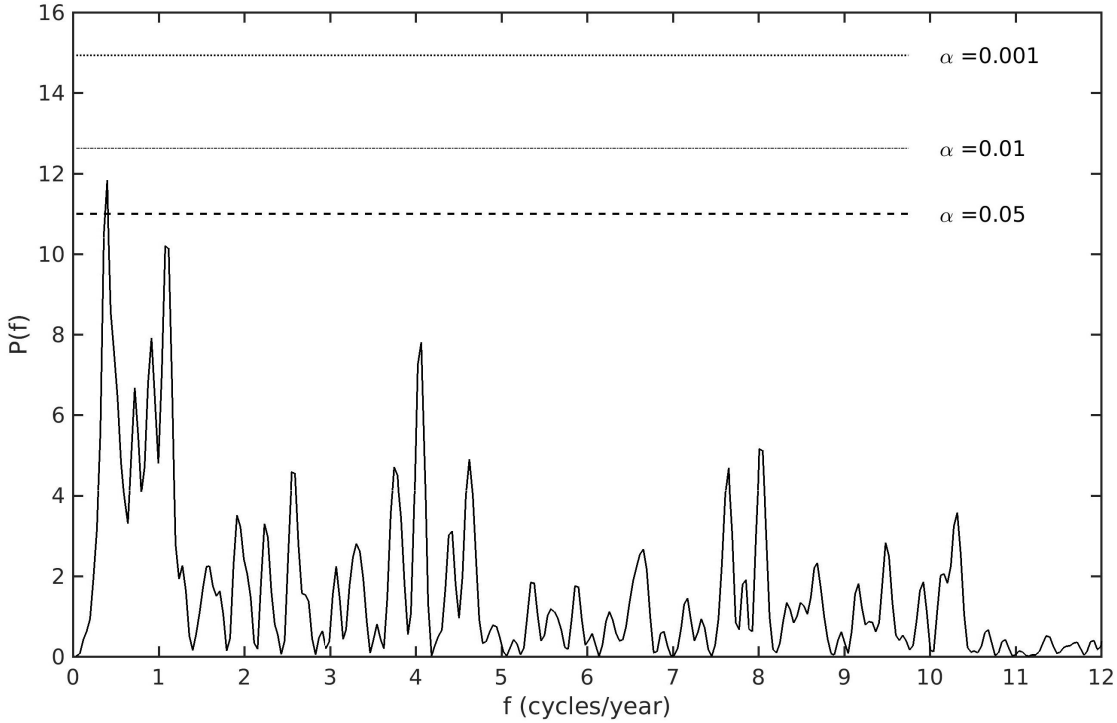

Figure 4.3: The same plot as in Figure 4.1 magnified to show the power spectrum in the low frequency range. A broad peak is evident from 0.4–1.0 cycles per year, with a significant spike at 0.4. A period of approximately one year may reflect a seasonal cycle in slopes.

(Table 4.1, Figure 4.4). The steeper slopes in winter, closer to α_{ff}, likely indicate an increase in fossil fuel influence during that season, when fossil fuel use remains steady or increases in the northeast due to energy use for heating while the biosphere is dormant and plant gas fluxes are small. This is consistent with the winter increase in CO_2 from fossil fuels at Harvard Forest observed by *Potosnak et al.* [1999]. Noise in the data may have obscured a stronger seasonal cycle, leading to the lack of a seasonal cycle in certain subsets of the data and periodicities varying from the expected value of one cycle/year.

Season	Mean	Median	SD	SEM	N
summer	-1.05	-1.05	0.22	0.01	1007
winter	-1.12	-1.15	0.44	0.02	441

Table 4.1: Average values for slopes categorized by season. Winter includes the months of December, January, and February; summer includes June, July, and August. Averages were calculated using the iteratively cut data set described earlier in Section 3.2. Significantly more negative slopes in winter may reflect increased fossil fuel influence. SD, standard deviation; SEM, standard error of the mean; N, number of data points.

Aside from a shifting balance between fossil fuels and the biosphere, a seasonal cycle in slope values could also be driven by cycles in plant physiological processes. Depending on the season, trees may produce different types of molecules with different stoichiometric ratios [*Seibt et al.*, 2004], although *Gallagher et al.* [2017] found that the oxidative ratio of leaves and tree rings was invariant throughout the year. Intense growth and photosynthesis in the summer contrast with reduced or absent photosynthesis during the winter, when fossil fuel influence also increases. Seasonal cycles in u* may increase the importance of local exchange during the summer and background variability during the winter, potentially contributing to a yearly cycle as well as the difference in winter vs. summer average slope values.

The three week cycle that appears in the day/high and night/low subsets is perplexing. Although potentially the result of some environmental factor, such as the passage of fronts, this cycle is possibly the result of an experimental factor, such as the

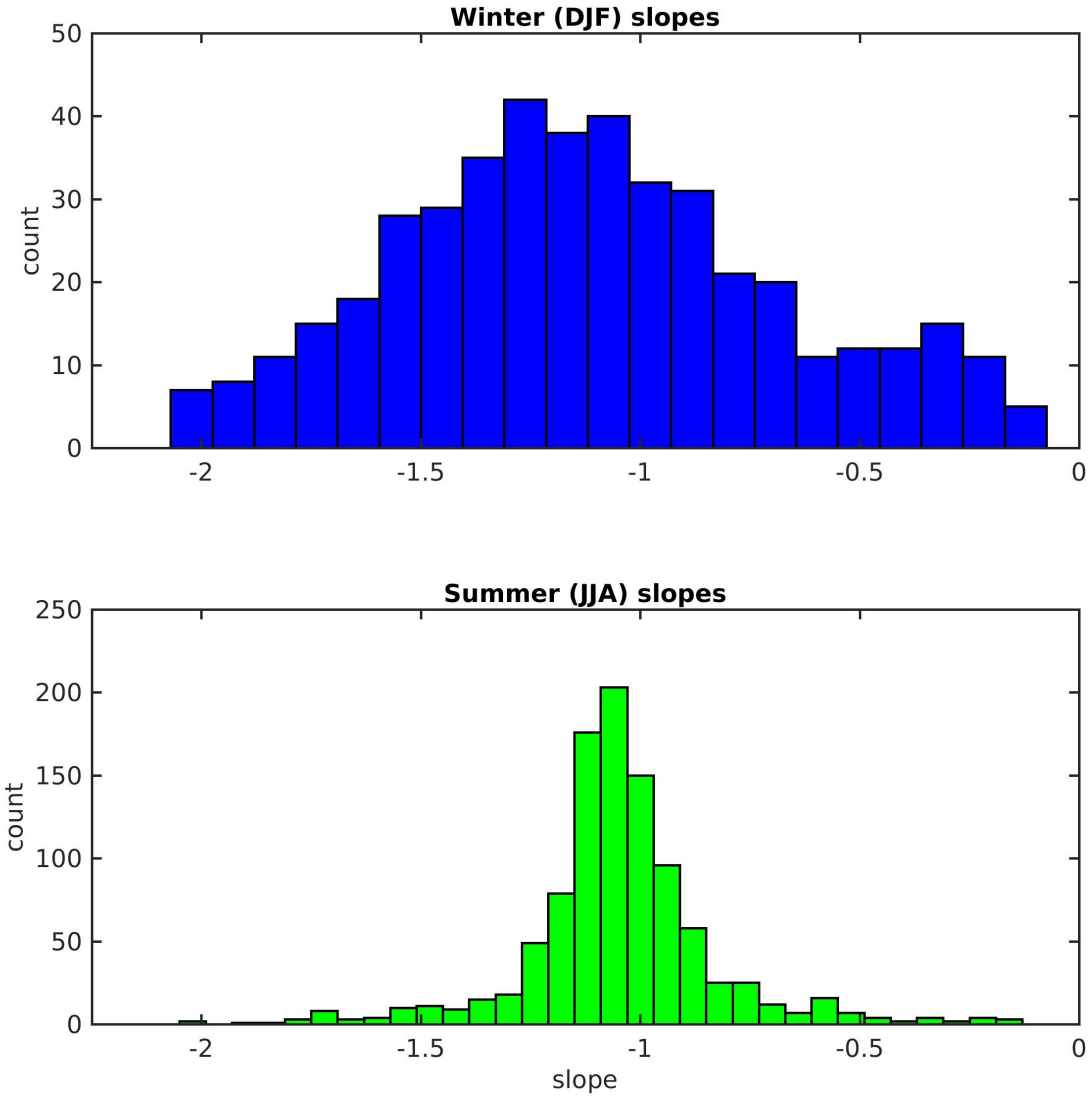

Figure 4.4: Histograms for winter and summer slope values. Winter comprises the months of December, January, and February and summer comprises June, July, and August. Winter slopes have a more negative mean value, closer to the stoichiometric ratio for fossil fuel combustion, although the sample size is smaller for winter than for summer. Statistics are shown in Table 4.1.

periodic runs of calibration tanks. One of the calibration tanks (the low frequency tank) is run approximately every three weeks, potentially producing a three-week cycle in the periodograms. This cycle might be more apparent in certain subsets of data if the low-frequency tank calibration runs were to occur, by chance, more often during day/high and night/low periods. This possibility will be explored in future work.

4.2 Analysis of environmental variables

To investigate potential drivers of the one year seasonal cycle observed in the slope data, we performed spectral analysis on a range of variables measured at the Harvard Forest EMS tower, including air temperature, net ecosystem exchange (NEE), ecosystem respiration, photosynthetically active radiation (PAR), and friction velocity (u*) [*Munger*, 2018]. All of these variables show strong, highly significant periodicities at seasonal (1 cycle/year) and daily (365 cycles/year) frequencies, with harmonics appearing at multiples of these frequencies. The yearly cycle corresponds with the yearly cycle observed in the slope data. In contrast, these environmental variables do not produce a peak at a period of three weeks, further indication that this cycle in the slope values may be an artifact of the sampling and calibration procedures.

Chapter 5

Modeling air transport and exchange

5.1 Introduction

A central concern in the preceding analysis is the link between our average slope value and the average $O_2 : CO_2$ exchange ratio for the terrestrial biosphere, α_b. After deriving a representative average for the stoichiometric ratio from our 6-hour plots, we must evaluate whether this slope provides an accurate measurement of α_b or simply reflects variability in the "background" composition. Only a local influence will give a representative value of α_b for a temperate forest; background signals reflect a mix of fossil fuel and biosphere influence from across the country which will vary with changes in the source region of air. The balance of local and background signals that we observe depends on both the size of the signals and on the dynamics of the atmosphere and boundary layer.

Ideally, we would subtract the variability of the background air upwind of Harvard Forest from our observations to extract a local signal. However, measurements of this background air do not exist, nor do we know what slope the background variability would have. Therefore, we use a different metric of background influence: the range of mixing ratios of both CO_2 and O_2 over 6-hour intervals. Variability driven by changes in the background (i.e. changes in the source region of the air) will be superimposed on local

exchange, so that a range in CO_2 and O_2 larger than that produced by local exchange alone must be driven by background variability in addition to local influence.

In the following chapter, we first use a Lagrangian model of atmospheric transport to investigate variability in the source region for air arriving at Harvard Forest. We then address the local vs. background question by considering whether a simple model of purely local fluxes can explain the full range of CO_2 and O_2 variability that we observe in our 6-hour intervals. If so, this increases our confidence in our interpretation of slopes as indicators of α_b. We also use the model to understand the relationship between surface fluxes, atmospheric transport, and predicted slopes.

5.2 Modeling of atmospheric transport

To examine variability in the source region and the geographic path of air arriving at Harvard Forest over the course of a 6-hour period, we used a Lagrangian model of atmospheric transport. HYSPLIT is an online atmospheric transport and dispersion model available on the National Oceanic and Atmospheric Administration (NOAA) website [*NOAA*, 2018]. For a specified location and sampling time, the model uses archived, gridded observations of winds, temperature, and pressure to calculate the past trajectory of an air parcel, which represents a set of properties for a given air mass [*Stein et al.*, 2015]. We used the model to create back-trajectories for air parcels arriving at Harvard Forest over a range of dates and times, during winter and summer as well as daytime and nighttime periods.

Using multiple sequential trajectories started every hour, we observed changes in the source region of Harvard Forest air over time. The source region for our study site typically lies to the west, including parts of southern Vermont, western Massachusetts, New York, and Pennsylvania. This occasionally varies to include Boston, Maine, and even the southern coastal region for longer time periods (Figure 5.1). The distance

traveled in a 6-hour period varies from approximately 50 to 200 km for the sample of 6-hour trajectories we created (Figure 5.2). Not surprisingly, longer periods show parcels traveling greater distances (compare to Figure 5.1). Limiting the time period analyzed to six hours therefore appears to restrict the source region to a modest, fairly homogeneous area. This source region may include small urban centers (such as Albany, Springfield, and on occasion, Boston) but for the most part consists of relatively rural, forested regions.

Although the source region does vary over time, even for a single six hour period, changes tend to be small. Seasonal variability in wind speed and the passage of fronts likely cause greater changes in this source region on certain days (Figures 5.3 and 5.4). To understand the relative importance of this source region variability, we need to understand the range of variability in CO_2 and O_2 potentially driven by local exchange. If the range of variability that we observe can be explained by local influences alone, our 6-hour slope values should reflect a signal dominated by the terrestrial biosphere, with some contamination from fossil fuel signal. While multi-day aggregation of data would provide greater statistical power, *Gerbig et al.* [2006] show that the region of influence grows rapidly to continental scales with increasing time periods, defeating our efforts to characterize a biospheric signal.

5.3 Investigation of CO_2 and O_2 variability using a 1-D box model

We created a simple one-dimensional box model to better understand the importance of local influence vs. background variation in our slope plots. We consider parcels of air starting at the beginning of a hypothetical 6-hour trajectory moving over a simple landscape, and we record the CO_2 and O_2 values that result at the endpoint of this trajectory

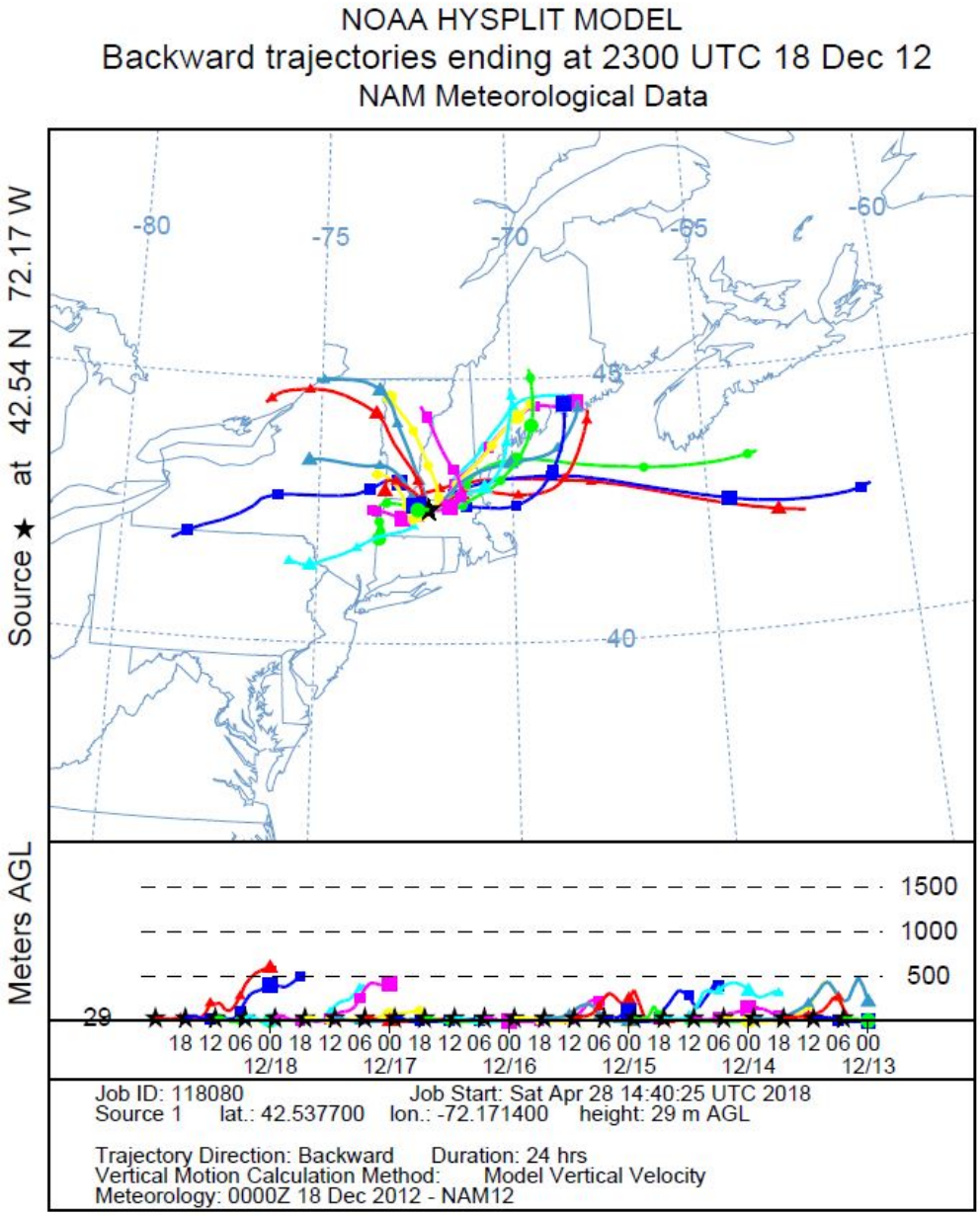

Figure 5.1: 24-hour back-trajectories for Harvard Forest produced by the HYSPLIT model, starting on 18 December 2012. Each trajectory shows the path of an air parcel during the 24 hours before it arrived at Harvard Forest. Trajectories were started at 6-hour intervals, moving back in time, to create the ensemble of trajectories shown. For this longer time period, the source region is highly variable and encompasses a large area, demonstrating our rationale for using a shorter 6-hour interval for our slope plots to reduce source region variability.

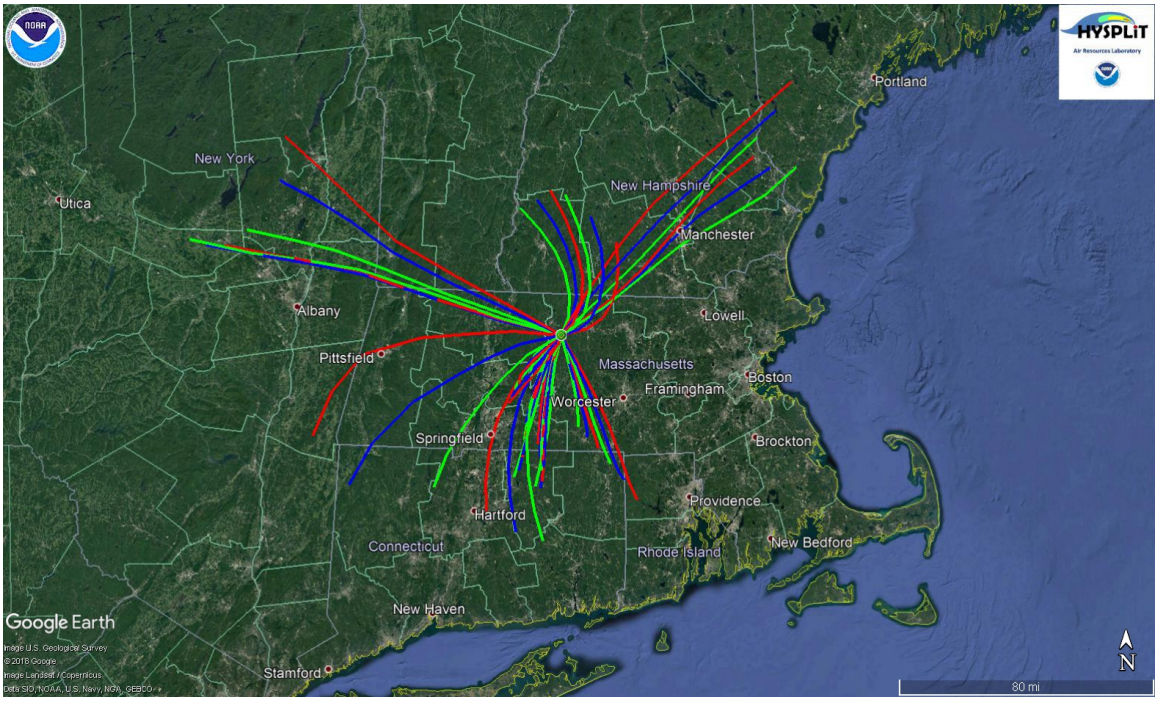

Figure 5.2: A set of 48 6-hour back-trajectories produced by the Hysplit model, plotted on satellite imagery from Google Earth. Each trajectory shows the path of an air parcel during the 6-hour period before it arrives at Harvard Forest. Trajectories include 24 daytime and 24 nighttime trajectories for two days in winter and two days in summer. For each day, trajectories were started each hour between 9am–3pm and 10pm–4am, the same time intervals used for the slope plots. These 6-hour intervals encompass a fairly consistent source region with a radius of 50–200 km. The area is mostly forested with some small urban centers.

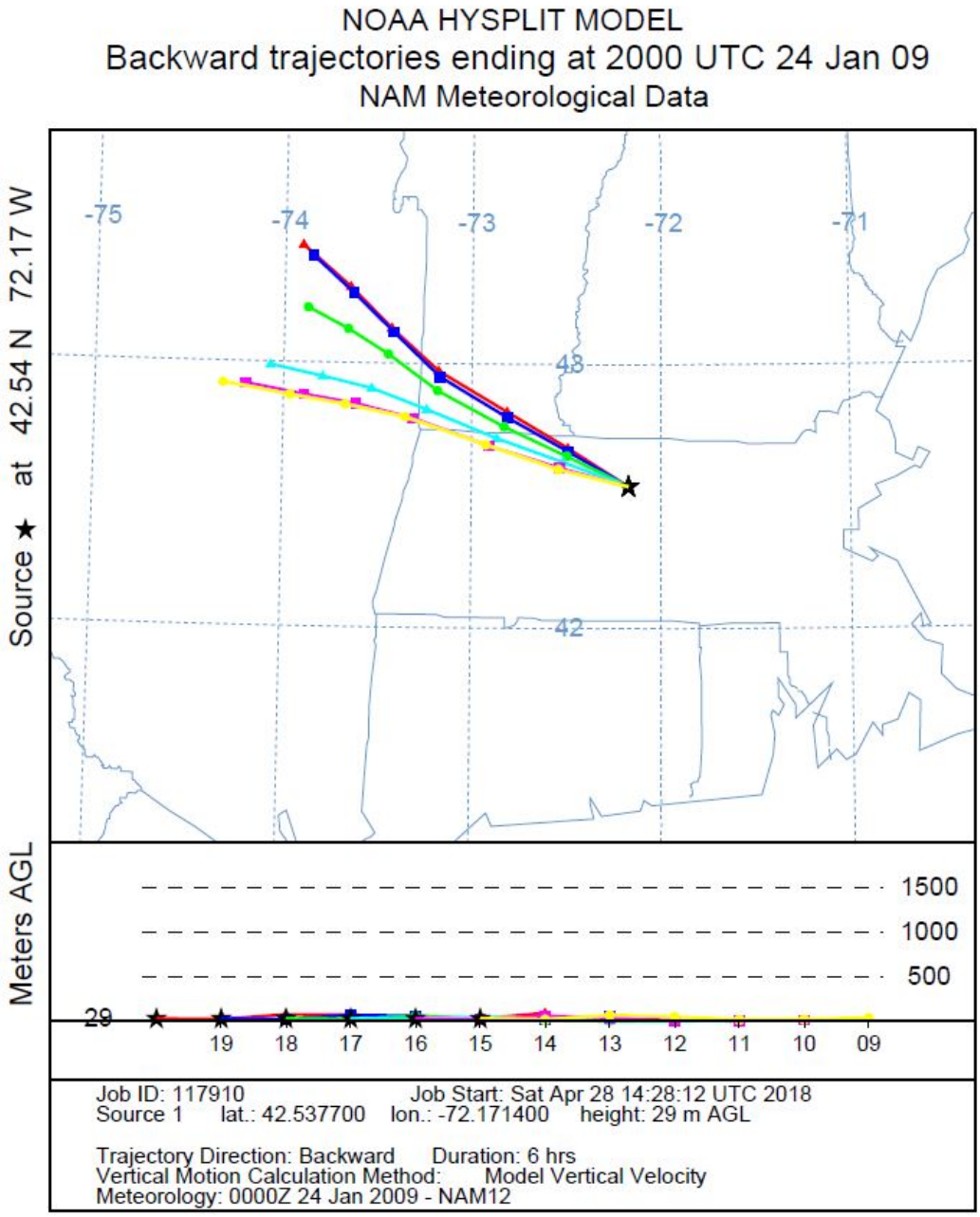

Figure 5.3: 6-hour back trajectories for Harvard Forest on 24 Jan 2009, with one trajectory started each hour between 9am and 3pm EST (corresponding to the 6-hour daytime interval used to calculate slopes), showing variability in the source region over that time period. This particular day shows a steady shift in the source region over time, although the path length and general source area remain similar. For this period, background variability might have a slightly stronger influence on the observed CO_2 and O_2 variability at Harvard Forest.

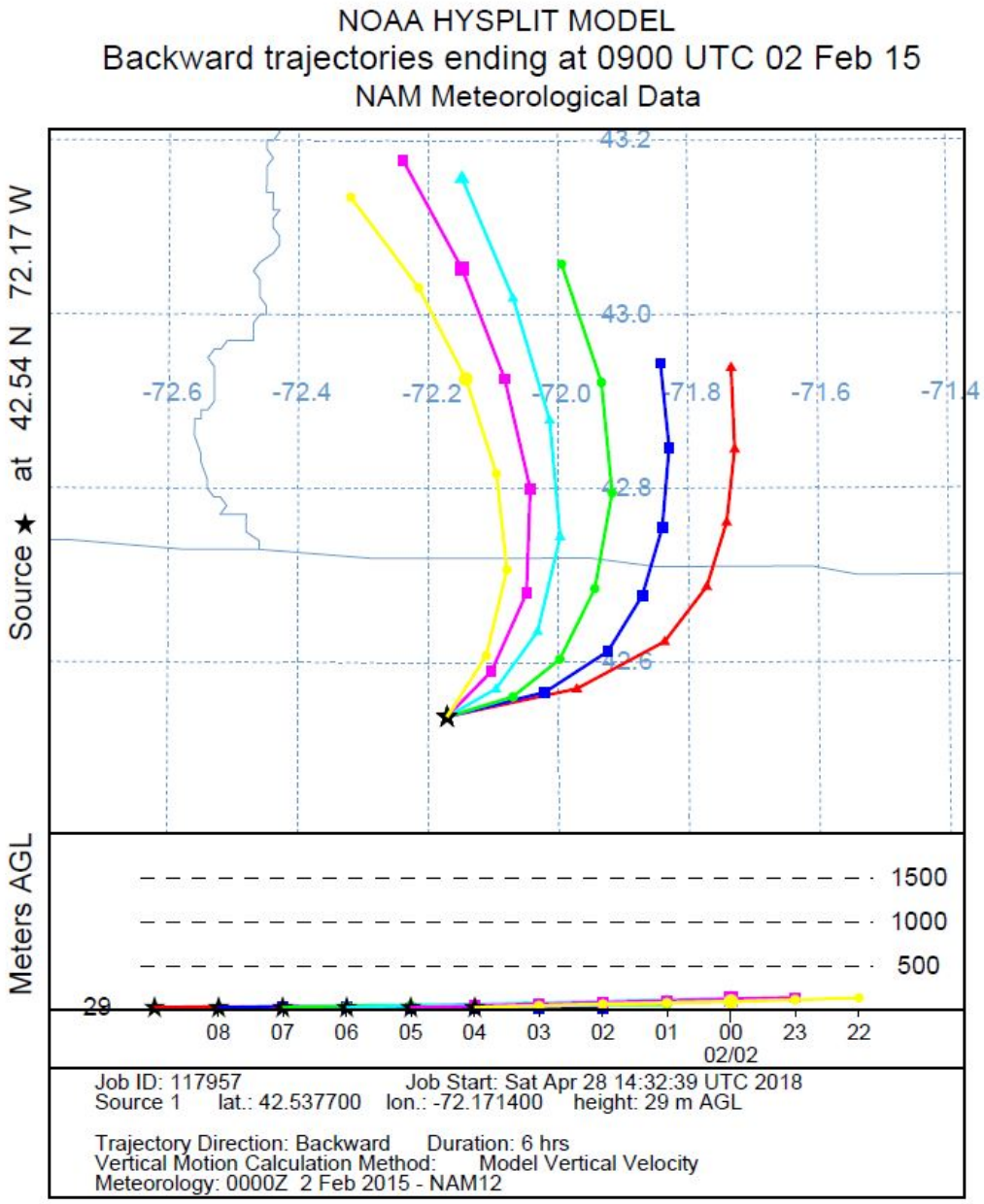

Figure 5.4: 6-hour back trajectories for Harvard Forest on 2 Feb 2015, with one trajectory started each hour between 10pm and 4am EST (corresponding to the 6-hour nighttime interval used to calculate slopes), showing variability in the source region over that time period. For this particular night, changes in the source region were minimal, with air coming from a small area in southwestern NH, so local influence would likely dominate the observed CO_2 and O_2 variability at Harvard Forest.

over time. By providing a sense of the range of CO_2 and O_2 variability produced by local exchange, this model helps distinguish between local influence and background variability. Rather than making quantitative predictions of CO_2 or O_2 for particular days, we seek to quantify the range of variability based on plausible estimates of surface fluxes and boundary layer heights.

The model calculates slope values resulting from a simplified scenario of land-atmosphere exchange over 6-hour periods. Each air parcel enters the first box of the model with a given "source" concentration of O_2 and CO_2. The model has 30 boxes, each representing the location of the air parcel at 12-minute intervals during the 6 hours it takes to travel from the first box of the model to the last box, when the air parcel arrives at Harvard Forest (Figure 5.5). As each parcel passes through each box, it exchanges O_2 and CO_2 with the landscape underneath that box. Any O_2 or CO_2 emitted or taken up by the landscape is added to or removed from the box volume, changing the mixing ratio of O_2 and CO_2 in the box based on the volume of the box and the magnitude of the flux using the relationship:

$$C_{final} = C_{initial} + \frac{F \cdot t}{h} \frac{22.4}{1000} \tag{5.1}$$

where C is the mixing ratio of O_2 or CO_2 in ppm, F is the flux from the landscape underneath the box in μmol/m^2/s, t is the length of the time interval over which the flux occurs in seconds, h is the height of the box in m, and the factors of 22.4 (the molar volume) and 1000 convert moles to liters and liters to m^3, respectively. We use one half of the PBL height as the box height, based on the value used in more sophisticated models of atmospheric transport [Maryann Sargent, personal communication; *Lin et al.*, 2003]. The O_2 and CO_2 values for each parcel change due to the fluxes from the landscape type beneath each box. We then shift the parcels forward by one box. After 24 hours of spin-up time to allow several parcels to pass through the entire sequence of boxes, the

O_2 and CO_2 concentrations of parcels are recorded as they arrive in the final "Harvard Forest" box at the end of the 30-box sequence. The time values are also recorded and slopes calculated separately for the same 6-hour daytime and nighttime intervals as the observed data to allow for comparisons in the range of CO_2 and O_2 variability over time.

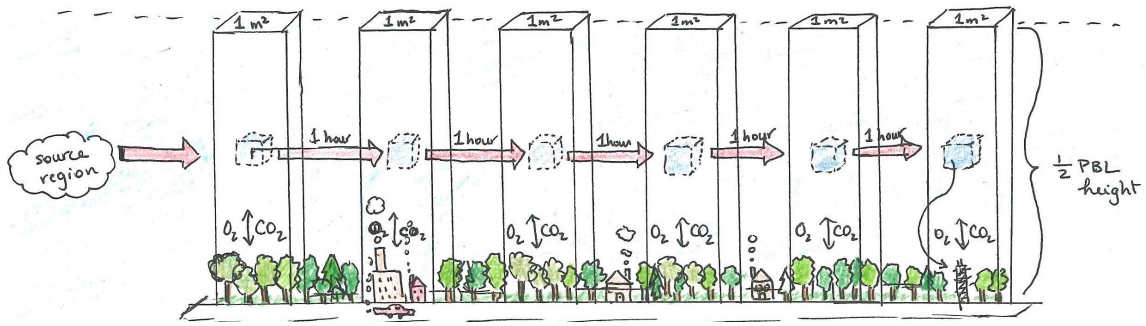

Figure 5.5: A conceptual diagram of the box model. Only 6 of the 30 boxes are shown, denoting the locations of the air parcel at each hour along its path to Harvard Forest. Air enters the model from the left (the source region) and leaves the model on the right (at Harvard Forest). As the air parcels travel through each box, O_2 and CO_2 are exchanged between the boxes and the landscape, which changes the mixing ratio of the parcel. The parcels shift forward every 12 minutes, and their mixing ratios are recorded as they arrive in the final "Harvard Forest" box on the right.

To parametrize fluxes of CO_2 from a forested landscape, we used CO_2 flux data from the Harvard Forest EMS tower [*Munger*, 2018]. Forests have negative CO_2 flux around noon, indicating uptake due to photosynthesis, and positive fluxes at night due to respiration. The hourly averages are based on 24 years of eddy flux measurements. Because fluxes vary greatly between winter and summer, we calculated separate climatologies for summer (June, July, and August) and winter (December, January, and February). We assumed an O_2 cycle of opposite sign with 1.05 times the magnitude of the CO_2 cycle, based on the most recent values of α_b from the literature [*Worrall et al.*, 2013] as well as our results. These daily cycles are shown in Figure 5.6.

To parametrize fossil fuel CO_2 fluxes, we used published CO_2 flux records collected in a range of urban and suburban areas. CO_2 fluxes vary depending on the size

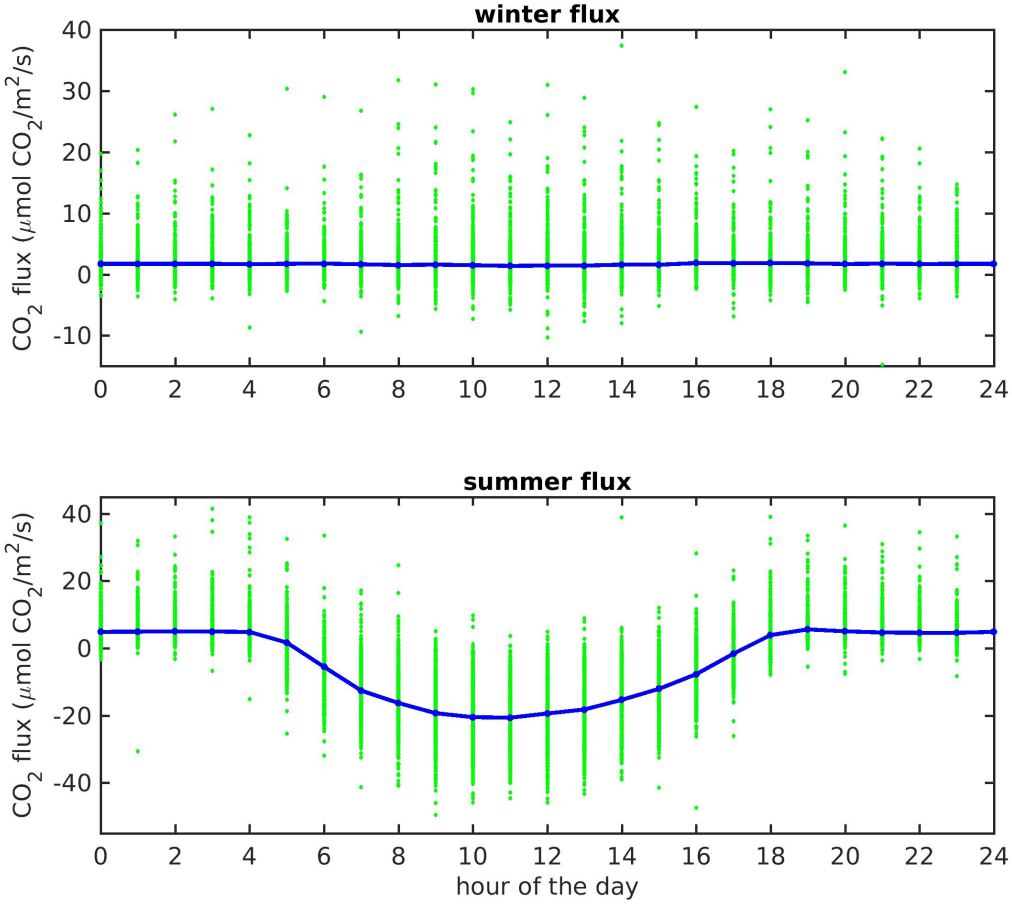

Figure 5.6: An hourly climatology of CO_2 fluxes at Harvard Forest based on hourly eddy flux data from 1991–2015, separated into winter and summer climatologies. Green dots show the ensemble of hourly data points, and the blue points and line show the average value for each hour. Summer fluxes have much higher magnitudes and a more pronounced daily cycle than winter fluxes, which are low, slightly positive (due to respiration) and fairly constant. One outlier in the summer plot (hour 23, flux 107.7 μmol $CO_2/m^2/s$) was omitted for clarity.

of the city and the area of green space, but tend to be positive at all times. CO_2 flux increases around 6am with the morning rush hour, followed by some drawdown at noon due to urban vegetation and a second peak around 6pm associated with the evening rush hour [*Bergeron and Strachan*, 2011; *Velasco and Roth*, 2010; *Ward et al.*, 2015]. Using a simplified model of this CO_2 cycle, we created a corresponding model of urban O_2 fluxes, assuming a stoichiometric ratio of 1.4 for fossil fuel combustion. For simplicity, we neglected the influence of urban vegetation on the stoichiometry. We also created CO_2 and O_2 cycles for a more suburban landscape, again based on eddy flux measurements taken in suburban areas [*Bergeron and Strachan*, 2011], with lower flux values overall due to reduced population density and increased green space. This may be more representative of the urban areas in the Harvard Forest footprint. The complete set of flux cycles used as input for the model is shown in Figure 5.7.

Along with the magnitude of flux from the landscape and the type of landscape, the height of the PBL determines the change in CO_2 and O_2 mixing ratios that results from a given flux. This height, which determines the volume into which surface O_2 and CO_2 fluxes are mixed, varies at several time scales, including a daily cycle, variation between days due to cloud cover, fronts, and storms, and seasonal variation with changes in day length [e.g. *Stull*, 2006]. We determined a representative daily profile for the PBL height using meteorological data (NAM 12 km) downloaded from the NOAA Air Resources Laboratory (ARL) archives. The data set has a temporal resolution of 3 hours and a spatial resolution of 12 km. We looked at PBL height for a subsample of dates, including 36 days in winter and 36 days in summer, and calculated an hourly climatology separately for each season (Figure 5.8).

We ran the model for 24 hours after spin-up time, plotted the concentrations of O_2 and CO_2, and calculated the stoichiometric ratio for the same daytime and nighttime intervals as for the observed data. To determine the sensitivity of the model to input

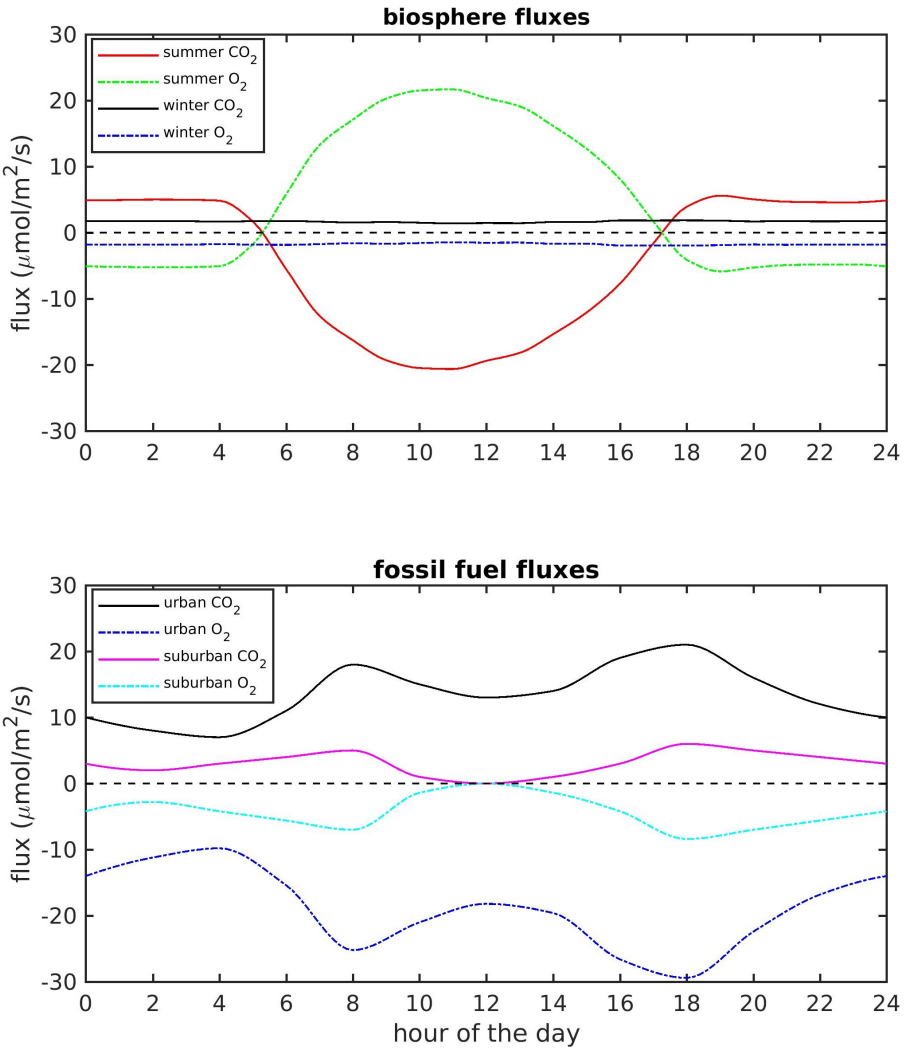

Figure 5.7: CO_2 and O_2 flux cycles used as input for the box model. Solid lines show CO_2 fluxes and dot-dash lines show corresponding O_2 fluxes calculated using the appropriate stoichiometric ratio. The top panel shows forest fluxes, separated into winter and summer cycles, and the bottom panel shows fossil fuel fluxes for an urban and suburban area. Although urban and suburban cycles do vary seasonally, the change is not as pronounced as the biospheric variability and is not essential for our simple model.

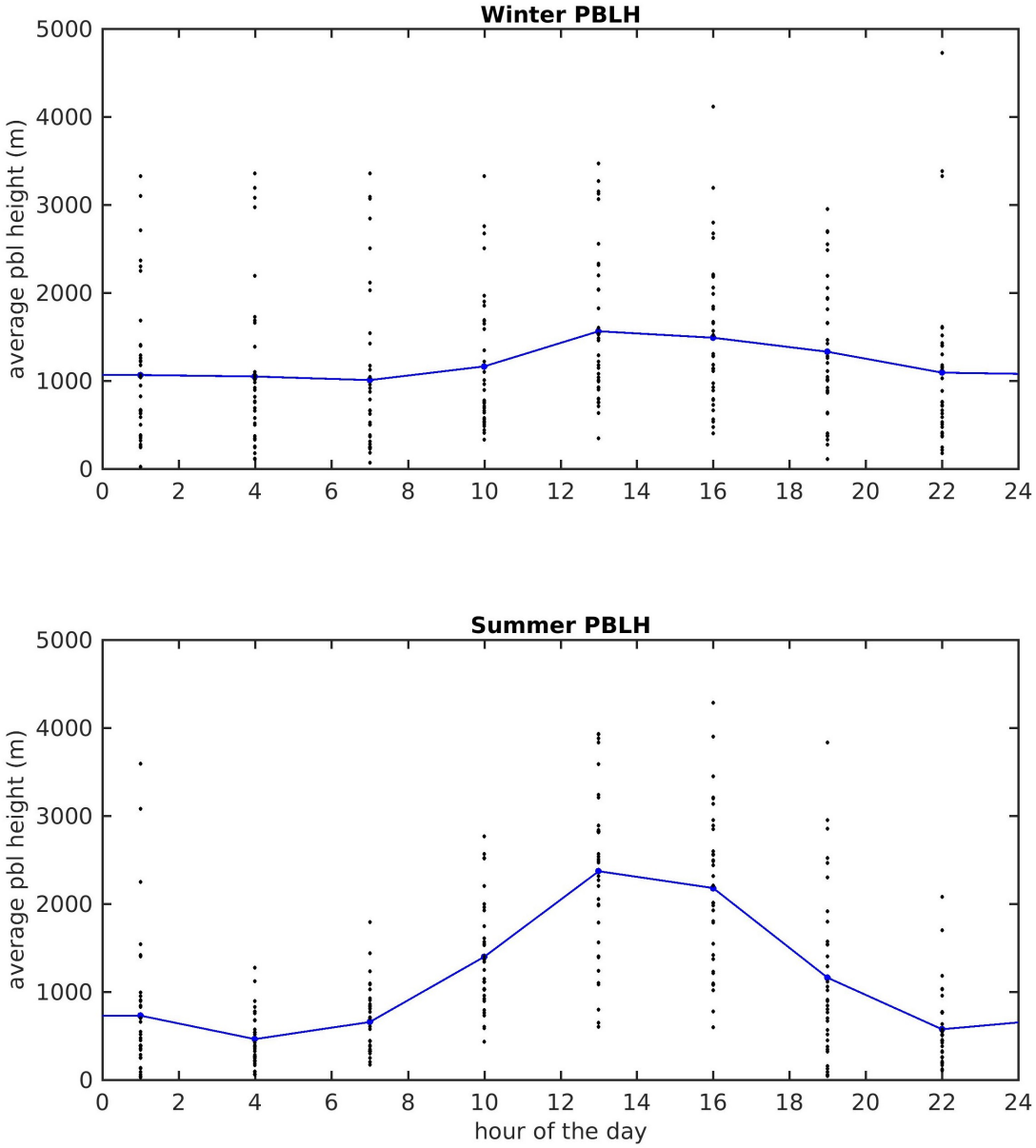

Figure 5.8: Daily cycles in PBL height for summer and winter calculated from ARL data [*NOAA*, 2017]. The black dots show hourly values (with a resolution of three hours) for 36 days during each season, and the blue line shows the average for each hour (with points every three hours connected by a linear interpolation). The summer PBL height is more variable, with lows in the morning and highs in the afternoon driven by radiative heating of the land, whereas the winter profile is relatively flat.

parameters, we varied the landscape composition, flux magnitude, and PBL height. As discussed above, the characteristics of the source region for air "entering" the model also impact the range of variability produced, but we used a constant source region to focus on the variability produced by local exchange alone in the 6-hour intervals.

5.4 Discussion of model results

There are three factors in our model that can cause slopes to differ from the input stoichiometric ratios: variability in landscape composition (changing fractions of urban and forest), variability in the fluxes for a given landscape type (e.g. winter forest vs. summer forest), and variability in the height of the PBL. Importantly, in addition to determining the measured slopes, these factors also determine the range of O_2 and CO_2 mixing ratios that the model produces.

For any PBL height and flux magnitude, a landscape of pure forest produces the expected slope value of -1.05 (Figure 5.9). A purely urban landscape produces the expected slope of -1.4 (Figure 5.10). We would expect mixed landscapes to produce intermediate slopes; however, in addition to the stoichiometry of the landscape fluxes, the slope depends on the relative magnitudes of the fluxes from urban and forested landscapes. Thus, a half-urban, half-forested landscape does not necessarily produce a slope halfway between -1.4 and -1.05. During the summer, when forest and urban fluxes are comparable, the slope is indeed intermediate (-1.33 at night, -1.18 during the day; Figure 5.11). However, during the winter when forest fluxes are consistently low, a small amount of urban influence can have a large effect on the slope value. Five boxes of urban landscape out of thirty, analogous to the passage of air parcels over a small city such as Albany or Springfield, produce a slope of -1.5, showing a strong fossil fuel influence (Figure 5.12).

In some cases, including the last scenario described above, the slope is actually outside the range of the two input stoichiometries. This results from the combined effects of the daily cycles in fluxes and the use of a constant source term for the model. For the winter scenario depicted in Figure 5.12, a parcel of air moving through the early-morning landscape arrives at Harvard Forest considerably enriched in CO_2 and depleted in O_2, relative to the "source region" concentration, because of the peak in urban CO_2 fluxes in the morning. During these early-morning hours, points on the plot evolve down-to-the-right with a slope near -1.4. As the morning progresses, parcels pass over the landscape at a time of smaller urban fluxes and larger forest fluxes. Points on the plot thus evolve up-to-the-left with a slope near -1.05. Taken over the entire interval, with no chronological infomation, the points have curvature (or simply scatter) and a straight-line fit yields a slope outside the expected range. This is one particular scenario that leads to seemingly implausible slopes, but similar arguments apply in different seasons and at different times of day. In short, real processes can lead to slopes steeper than fossil fuel combustion or shallower than forest exchange. They do not point to problems with the box model.

As discussed in the introduction to this chapter, our primary objective for the model is to characterize the range of CO_2 and O_2 variability due to local sources. With this in mind, we note that changes in the magnitude of the forest and urban fluxes as well as the PBL height also influence the range of CO_2 and O_2 variability. For example, summer PBL values tend to generate larger ranges in CO_2 and O_2 than winter values, because the top of the PBL dips to lower heights at night and in the early morning, amplifying the effects of surface fluxes at those times. This nighttime amplification turns out to be more important than the flux dilution that occurs on summer afternoons due to a relatively high PBL. Because of PBL dynamics, as well as daily cycles in surface fluxes, CO_2 and O_2 ranges during the daytime interval tend to be larger than those during the nighttime interval.

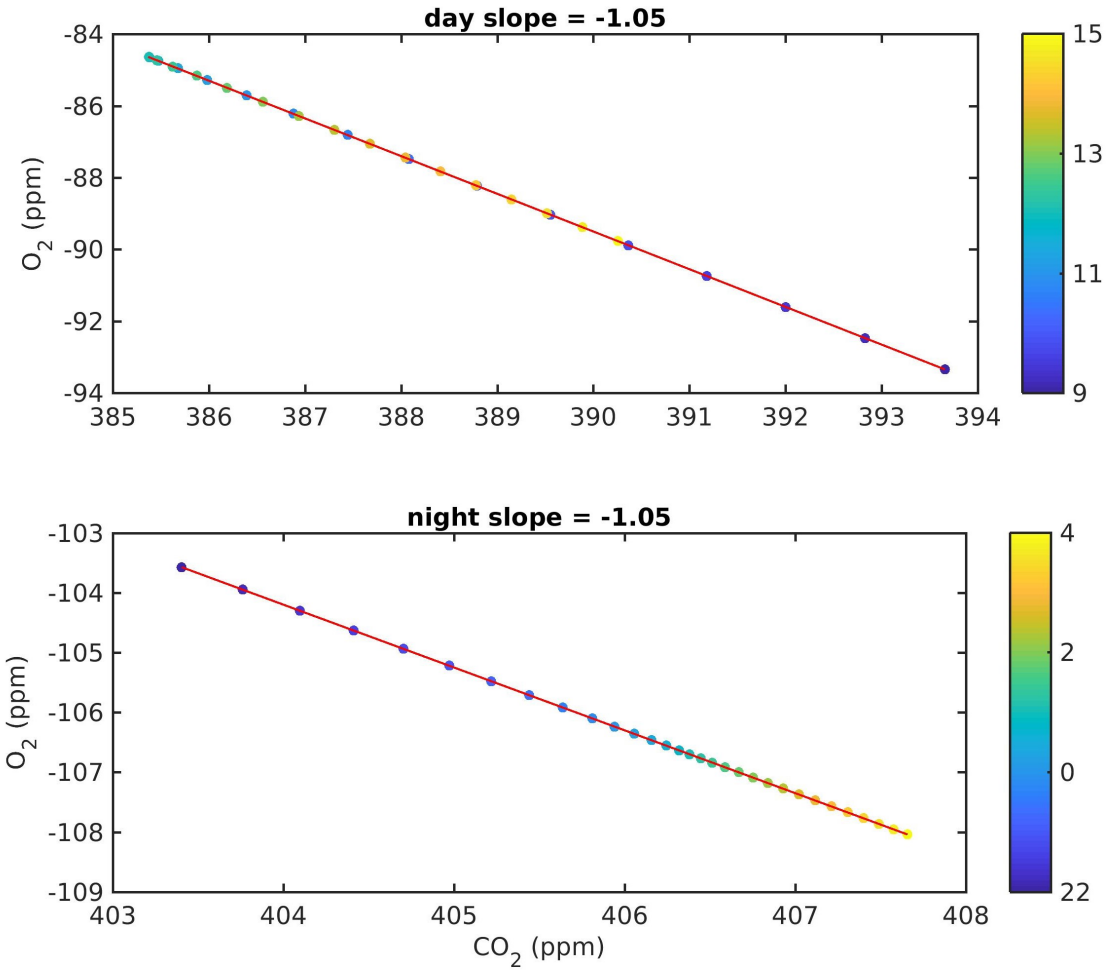

Figure 5.9: Model output for a forested landscape, showing the resulting scatterplots of O_2 vs. CO_2 for the same daytime and nighttime intervals used with the observed data. The color of the points indicates the time each arrived at Harvard Forest, with the labels on the colorbar denoting hours of the day. Note the decrease followed by an increase in CO_2 during the daytime interval and the consistent increase in CO_2 during the nighttime interval. The red line shows the linear regression associated with the slope value. This plot was produced using the summer PBL height and summer forest flux cycle. A completely forested landscape produces the expected slope of -1.05, reflecting the stoichiometric ratio of the forest fluxes used to drive the model.

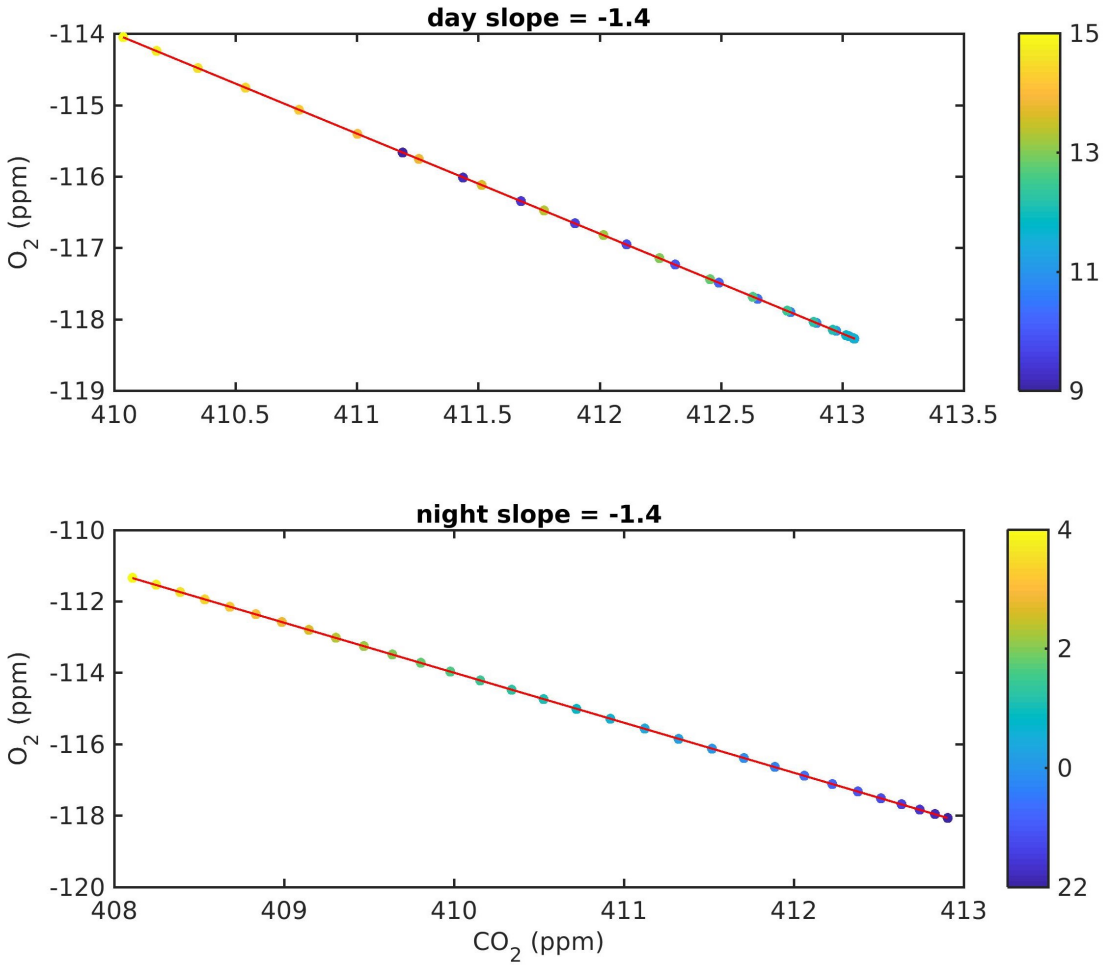

Figure 5.10: Model output for an urban landscape, using the winter PBL height and urban flux cycle. A completely urban landscape produces the expected slope of -1.4.

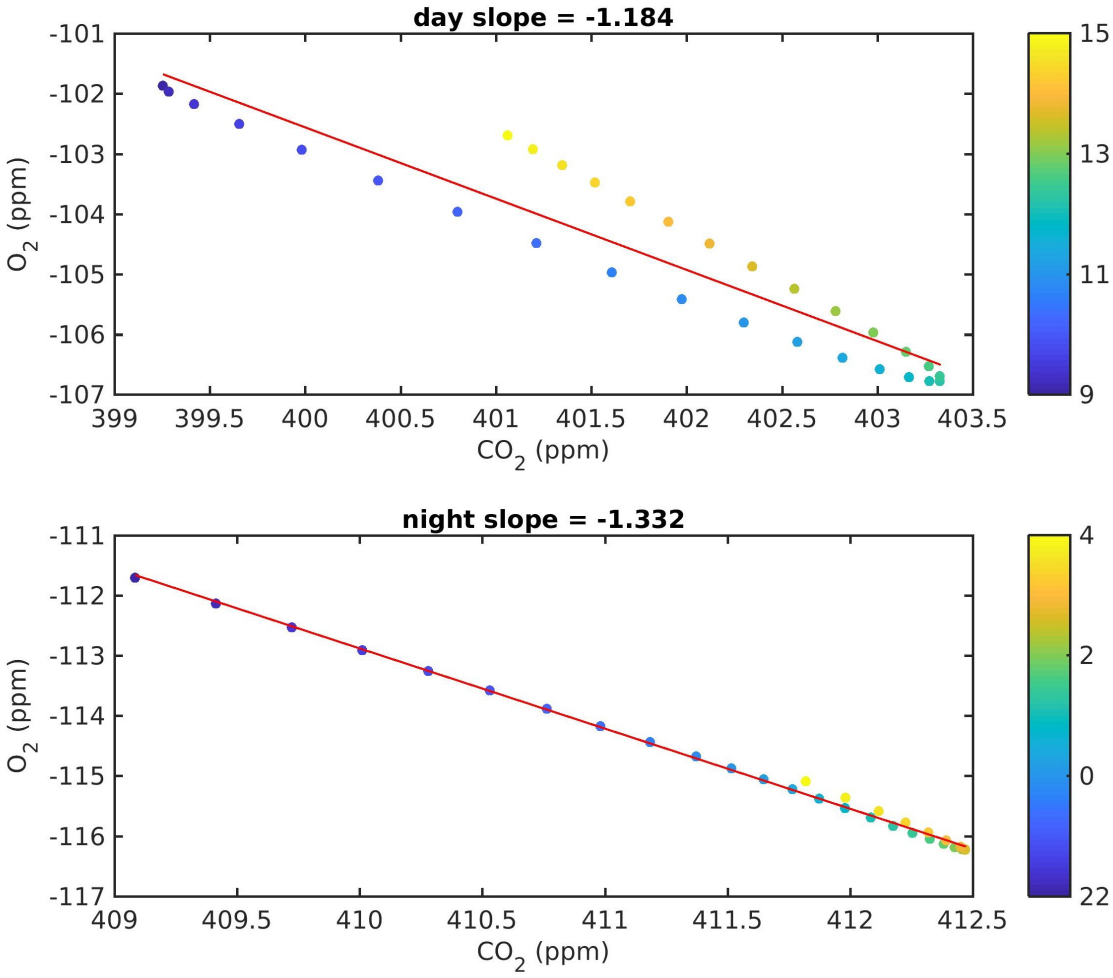

Figure 5.11: Model output for a mixed landscape consisting of 15 boxes of urban landscape followed by 15 boxes of summer forest, using the summer PBL height cycle. Because the summer forest fluxes are of similar magnitude to the urban fluxes, the slope produced is intermediate between -1.05 and -1.4, the input stoichiometries of the two landscape types.

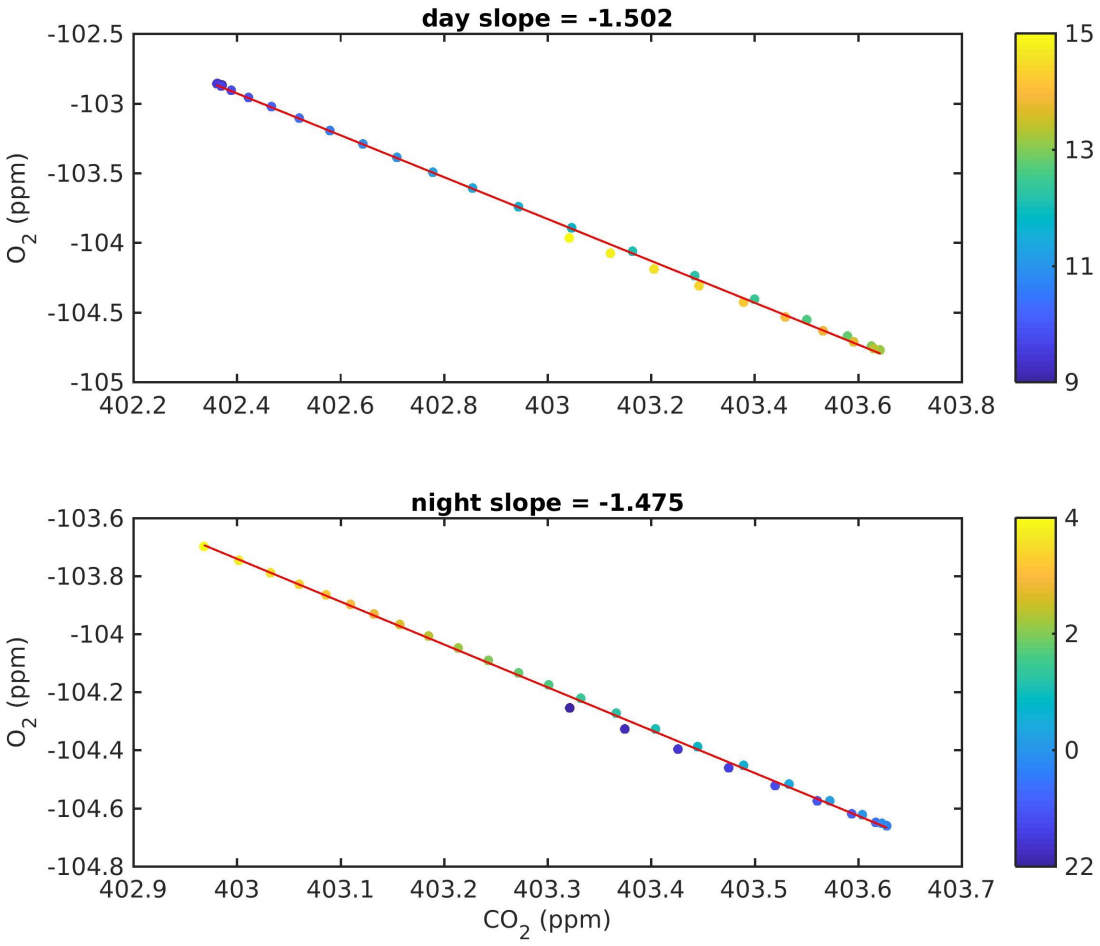

Figure 5.12: Model output for a mixed landscape consisting of 5 boxes of urban landscape followed by 25 boxes of winter forest, using the winter PBL height cycle. The resulting slopes are outside the range of the two input stoichiometries (-1.05 to -1.4), because the use of a constant source term for the model results in alternating increases and decreases in CO_2 and O_2 throughout the day based on the daily flux cycles, producing point patterns outside of a straight line and slope values outside the range of the inputs.

The model output is quite sensitive to the PBL height; a 50% reduction in PBL height causes a 100% increase in CO_2 and O_2 ranges, while a 150% increase in PBL causes a 30% reduction in the ranges. The magnitude of fluxes has a similar impact on observed CO_2 and O_2 ranges, with increased fluxes leading to larger ranges and vice versa. The phasing of PBL height relative to the flux cycles is also important. A 5-hour shift in the PBL cycle produces unphysical plots with concurrent increases in CO_2 and O_2 because of the unnatural offset between the flux cycles and the PBL cycles.

By comparing the range of CO_2 and O_2 predicted by the model with the ranges we observe in our slope data, we can determine whether the observed range of variability can be attributed to local influence alone or whether an additional contribution from background variability is required. Results of multiple model runs and sensitivity tests are summarized in Table 5.1. For summer conditions, the model produces CO_2 variability of about 4–16 ppm and O_2 variability of about 5–20 ppm (although the range for both is sometimes as low as 2 ppm at night). Winter ranges are smaller, on the order of 1–5 ppm for CO_2 and 1–6 ppm for O_2. If observed ranges of CO_2 and O_2 are similar to those produced by the model, this supports the conclusion that local influence drives the observed changes in mixing ratios; if the observed ranges are larger, this suggests an additional contribution from background variability.

For comparison with observed data, we recorded and analyzed the range of CO_2 and O_2 for each of the 3760 slope plots. For each plot, we calculated the range between the 95^{th} and 5^{th} percentiles for the CO_2 and O_2 values in each 6-hour interval. By using percentiles, we hoped to reduce the impact of outlier data points on the resulting ranges. However, because of some plots with anomalously high CO_2 and O_2 data points, the distribution of ranges includes some large values and is strongly right-skewed (Figure 5.13). We therefore used the median as a measure of central tendency. The median CO_2 range is about 8 ppm, and the median O_2 range is about 14 ppm.

Landscape	PBL	Day			Night		
		slope	CO_2 range	O_2 range	slope	CO_2 range	O_2 range
All summer forest	0.5×summer	-1.05	8.3	8.7	-1.05	4.3	4.5
5 urban, 25 summer forest	0.5×summer	-1.02	4.8	4.8	-1.19	4.1	4.8
15 urban, 15 summer forest	0.5×summer	-1.18	4.1	4.9	-1.33	3.4	4.5
25 urban, 5 summer forest	0.5×summer	-1.48	6.9	10.2	-1.44	2.6	3.6
All urban	0.5×summer	-1.40	9.7	13.5	-1.40	2.7	3.7
5 urban, 10 summer forest, 5 suburban, 10 summer forest	0.5×summer	-1.00	4.7	4.9	-1.19	3.5	4.1
All winter forest	0.5×winter	-1.05	0.49	0.52	-1.05	0.23	0.24
5 urban, 25 winter forest	0.5×winter	-1.50	1.3	1.9	-1.48	0.66	0.96
15 urban, 15 winter forest	0.5×winter	-1.42	3.5	5.0	-1.41	2.3	3.3
All urban	0.5×winter	-1.40	3.0	4.2	-1.40	4.8	6.7
All summer forest	0.25×summer	-1.05	16.6	17.4	-1.05	8.5	8.9
All summer forest	0.75×summer	-1.05	5.5	5.8	-1.05	2.8	3.0
All 0.75×summer forest	0.5×summer	-1.05	6.2	6.5	-1.05	3.2	3.3
All 0.25×summer forest	0.5×summer	-1.05	10.4	10.9	-1.05	5.3	5.6

Table 5.1: Summary of CO_2 and O_2 ranges and slopes for daytime and nighttime intervals produced by the model for a range of landscape compositions, PBL heights, and flux magnitudes. Summer ranges tend to be larger than winter ranges, and daytime ranges tend to be slightly larger than nighttime ranges. Increases in the magnitude of the fluxes or decreases in the PBL height result in increased ranges.

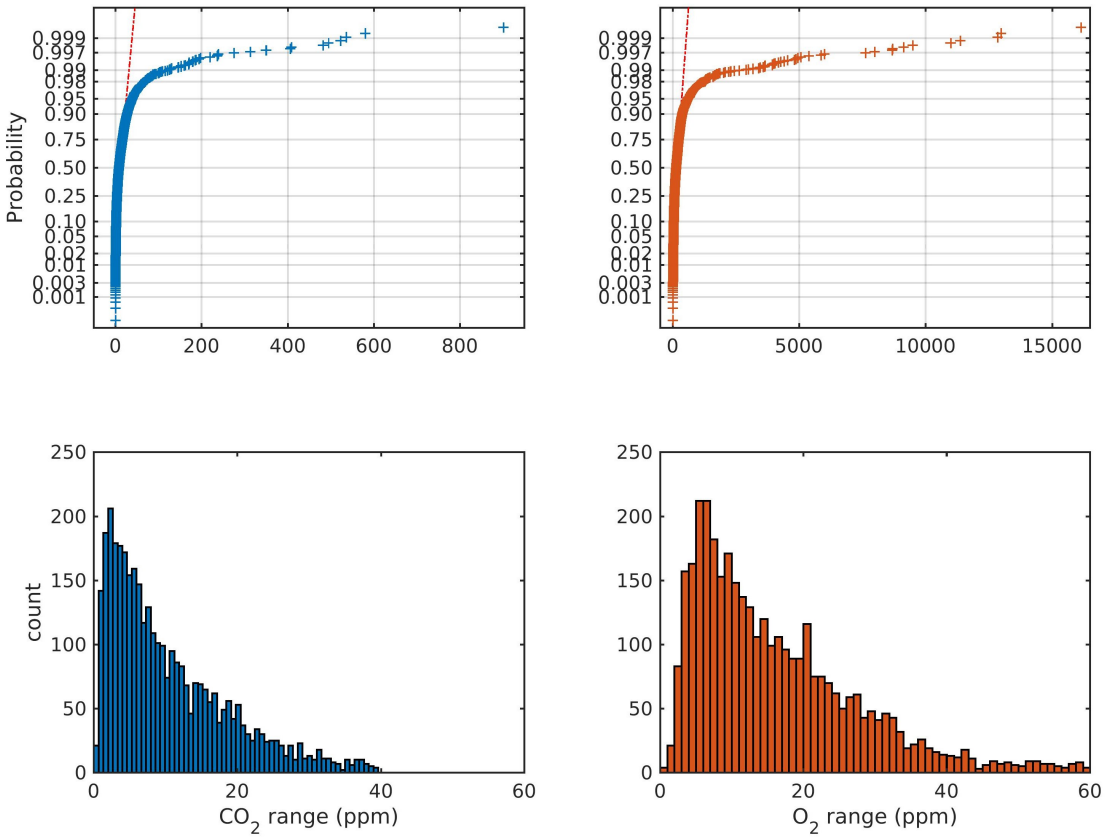

Figure 5.13: Histograms showing the distribution of CO_2 and O_2 ranges for the full set of 6-hour slope values. Top two panels show normal probability plots, where the dashed line describes a normal distribution. CO_2 ranges are slightly lower than O_2 ranges overall. One outlier in the normal probability plots with a CO_2 range of 15080 ppm and an O_2 range of 24760 μmol/mol was omitted for clarity, and the upper tail of the histograms was omitted to better show the distribution of the core of the data.

The observed CO_2 and O_2 variability is consistent with the variability produced by the model under a range of conditions. Although many of the CO_2 and O_2 ranges for individual 6-hour intervals are larger than those produced by the model, the distribution supports the conclusion that a substantial portion of our slopes reflect a strong local signal.

If the model more realistically included small changes in the concentration of "source region" air, this would only increase the modeled ranges, further supporting the notion that we do observe 6-hour intervals dominated by local influence at our site. Alternatively, if we were to limit our observations to periods with "stable" meteorology, we might well reduce the high side of the range and improve data-model agreement. The latter possibility is discussed further below.

Chapter 6

Isolating periods of local influence to better measure α_b

6.1 Introduction

Based on the model results, local exchange produces CO_2 and O_2 variability that is similar to the median ranges in our observed data. However, many of the slope values from the observed data have ranges larger than the values that characterize local influence, and these plots must therefore be driven by background variability as well as local exchange. In particular, many of our winter plots have CO_2 and O_2 ranges much larger than those predicted by local exchange alone. In order to more accurately determine the value of α_b, we attempt to remove those slopes influenced by changes in the source region and focus on slopes dominated by local exchange. For this analysis, we use the iteratively cut data set from which the averages described in Section 3.3 were derived.

6.2 Slopes categorized by friction velocity

To filter slopes dominated by background variability out of the dataset, we first used threshold values for u*. Friction velocity reflects turbulent motion in the air, so that large

values of u* are associated with more mixed conditions and low values are associated with more stratified conditions. We hypothesized that periods with low u* would be more likely to reflect local influences, and thus contain slope values that are closer to α_b. We used the same u* data from the Harvard Forest EMS tower referenced earlier [see Section 3.1; *Munger*, 2018] to match individual u* values to individual gas concentration data points. Each CO_2/O_2 data point was matched to a u* value using the hour of the time at which the CO_2/O_2 data point was collected. Then the average, maximum, and standard deviation in u* were calculated for each 6-hour interval, associating values for this set of parameters with each slope value. Maximum u* may reveal the occurrence of gusts that disturb canopy stratification, while σ_{u*} may similarly reflect instability that would not be evident from a simple average. Because u* tends to be higher during the winter, this filter may also emphasize summer slopes whose CO_2 and O_2 ranges reflect more of a local influence.

Efforts to categorize slopes based on u* were inconclusive. There was little difference between slopes split using thresholds of 40 cm/s for average u*, 50 cm/s for maximum u*, and 10 cm/s for σ_{u*}. Although we expected higher values of these parameters to be associated with steeper slopes, due to windy conditions increasing turbulent exchange with air outside the canopy layer and increasing the footprint for a given time period, in some cases the slopes with lower u* values had a slightly more negative mean. The high variability in u* within each daily period and its sensitivity to wind gusts may in fact weaken this parameter's ability to distinguish between local and background influence by being overly sensitive, or u* may simply not be closely related to the size of the tower's footprint for a given 6-hour period. A scatter plot of CO_2 ranges against average u* showed no significant positive relationship, indicating that u* may not provide a useful filter for selecting periods of local influence. Because the sensor for u* is located above the canopy, it may not reflect air movement dynamics at the low intake, which may further obscure any relationship between u* and slopes.

6.3 Slopes filtered by CO_2 range

Because CO_2 and O_2 ranges are also a metric of the amount of local influence vs. background variability, we filtered slopes by CO_2 range. We performed this process separately by season, because the CO_2 and O_2 ranges expected for local influence are much smaller during the winter. Results of this range-based filtering are shown in Table 6.1.

For the summer, slopes with low CO_2 ranges have a less negative mean than those with higher CO_2 ranges, using a threshold of 10 ppm. Although there are fewer slopes for the low range subset than the high range (176 vs. 831), the dataset for the low CO_2 range is still large enough to maintain statistical significance.

For the winter dataset, fewer slopes are available, but the slopes with lower CO_2 ranges still have a less negative average value than the slopes with higher CO_2 ranges, using a threshold of 4 ppm. The difference between the two averages is enhanced compared to the summer, despite a steeper average slope for the low-range subset compared to the summer low-range subset. We speculate that this steeper slope for the winter low-range subset is due to the higher likelihood of fossil fuel signal contaminating the observed $O_2 : CO_2$ exchange ratio. A potential increase in the influence of soil respiration in winter may also result in a different average exchange ratio than in the summer.

Season	Filter	mean	median	SD	SEM	N
Summer	CO_2 range < 10 ppm	-1.00	-1.00	0.30	0.02	176
	CO_2 range > 10 ppm	-1.06	-1.06	0.19	0.01	831
Winter	CO_2 range < 4 ppm	-1.06	-1.10	0.48	0.04	188
	CO_2 range > 4 ppm	-1.16	-1.18	0.39	0.02	253

Table 6.1: Slopes filtered by CO_2 range and season. For the summer, a range of 10 ppm or less in CO_2 should reflect local influence, based on our model results, whereas for winter, the threshold is lower, about 4 ppm. As expected, the average slope values for the low CO_2 range subsets were shallower than those for the high CO_2 range subsets. The summer slope values were also shallower than the winter values, perhaps an indication of fossil fuel influence during the winter. SD is standard deviation, SEM is standard error of the mean.

Chapter 7

Conclusion

Our results indicate that the use of atmospheric data to estimate the value of α_b is a promising technique that may complement elemental analysis techniques. Based on our average value for the exchange ratio, we suggest that the true value of α_b may be lower than the accepted value of 1.1, at least for a temperate deciduous forest ecosystem, with a potential range of 1.00 to 1.06. Some evidence of seasonal cycles as well as significantly different average slope values for winter and summer suggest that the value of the $O_2 : CO_2$ exchange ratio varies seasonally. Possible drivers include variation in biospheric processes, such as changes in the types of molecules produced by plants and the influence of soil respiration, as well as an increase in fossil fuel influence during the winter due to the reduced magnitude of biospheric fluxes. Based on the results of a simple one-dimensional box model, we estimated the range of CO_2 and O_2 variability that could plausibly result from local exchange alone, and we used this range to obtain a subset of slopes with CO_2 ranges indicative of a strong local signal. The average value of these slopes for the summer months, 1.00 ± 0.02, constitutes our best estimate for the value of α_b for a summer biosphere, the time when biospheric processes are most active. We speculate that our winter value of 1.06 may be contaminated by a fossil fuel signal and may not reflect a purely biospheric respiration signal.

The atmospheric measurement technique, although complex in interpretation, offers valuable insight for a seasonal- and ecosystem-averaged value of α_b. This average value is particularly useful for carbon sink calculations, which lump the entire terrestrial biosphere carbon uptake as a single term. In a previous study, *Seibt et al.* [2004] discounted the use of canopy air measurements to determine the $O_2 : CO_2$ exchange ratio for the terrestrial biosphere, due to canopy air consisting of mixed exchange ratios from plants, soil, and air introduced from outside the canopy by turbulent exchange. However, this earlier work did not include long time series measurements. Although background variation may influence some of our observed slopes, careful consideration of atmospheric conditions and CO_2 ranges allows for the selection of periods dominated by a local signal. In fact, the spatiotemporally-averaged nature of our value of α_b is one of its strengths. By encompassing all types of carbon assimilation and respiration for a particular ecosystem over time, our method avoids the need to quantify separate contributions from various carbon pools (trees, soil, understory plants, etc.) and from various processes within each pool (production of glucose, lignin, etc.) to the overall exchange ratio.

Further investigation into data filtering methods that isolate periods of local influence may enhance the applicability of concurrently-measured atmospheric O_2 and CO_2 data as a measurement of α_b. One technique may be to use slopes from stagnant summer periods with high pressure, when wind speeds are low, CO_2 and O_2 ranges are small, and local influence is likely dominant. Another potential technique could use the high intake, which tends to have higher CO_2 and O_2 variability, as a reference or background measurement, and look at the difference between the low intake and the high intake as an indicator of local influence at the low intake. Selecting a study site that is more isolated from fossil fuel activity may also provide a more accurate measurement of α_b, although such sites are not available for all ecosystem types. Use of concurrent measurements of CO as a tracer of fossil fuel signal could help filter out time intervals significantly influenced by fossil fuels [*Potosnak et al.*, 1999].

Recent research suggests that species type rather than climate has a strong influence on the $O_2 : CO_2$ exchange ratio in different ecosystems [*Gallagher et al.*, 2017]. Atmospheric measurements in a range of ecosystem types, from tropical forests to tundra, would therefore provide important information on global variability in α_b, allowing for the calculation of a composite value for α_b [as in *Worrall et al.*, 2013] that takes into account the known distributions of different ecosystem types. A more accurate value for α_b will improve our knowledge of the land carbon sink and forecasts for future atmospheric CO_2 levels. In our rapidly changing climate and world, this knowledge is critical for both mitigation and adaptation efforts.

Appendix A

Black & White Figures

Figure A.1: Map showing the location of the Harvard Forest EMS tower and Bowdoin College (Figure 2.1).

Figure A.2: Photograph of the aspirated inlet at the high intake, designed to reduce thermal fractionation. Image credit: Mark Battle (Figure 2.2).

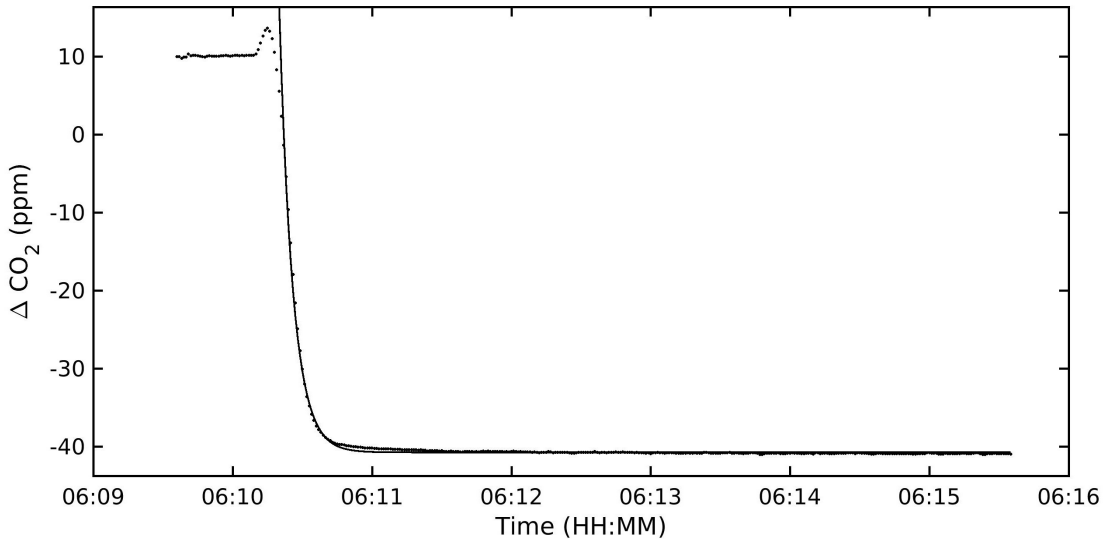

Figure A.3: Plot of CO_2 vs. time for a calibration run, showing the curve fitted to the data points about which the scatter is calculated to determine the instrumental error in CO_2 measurements (Figure 2.4).

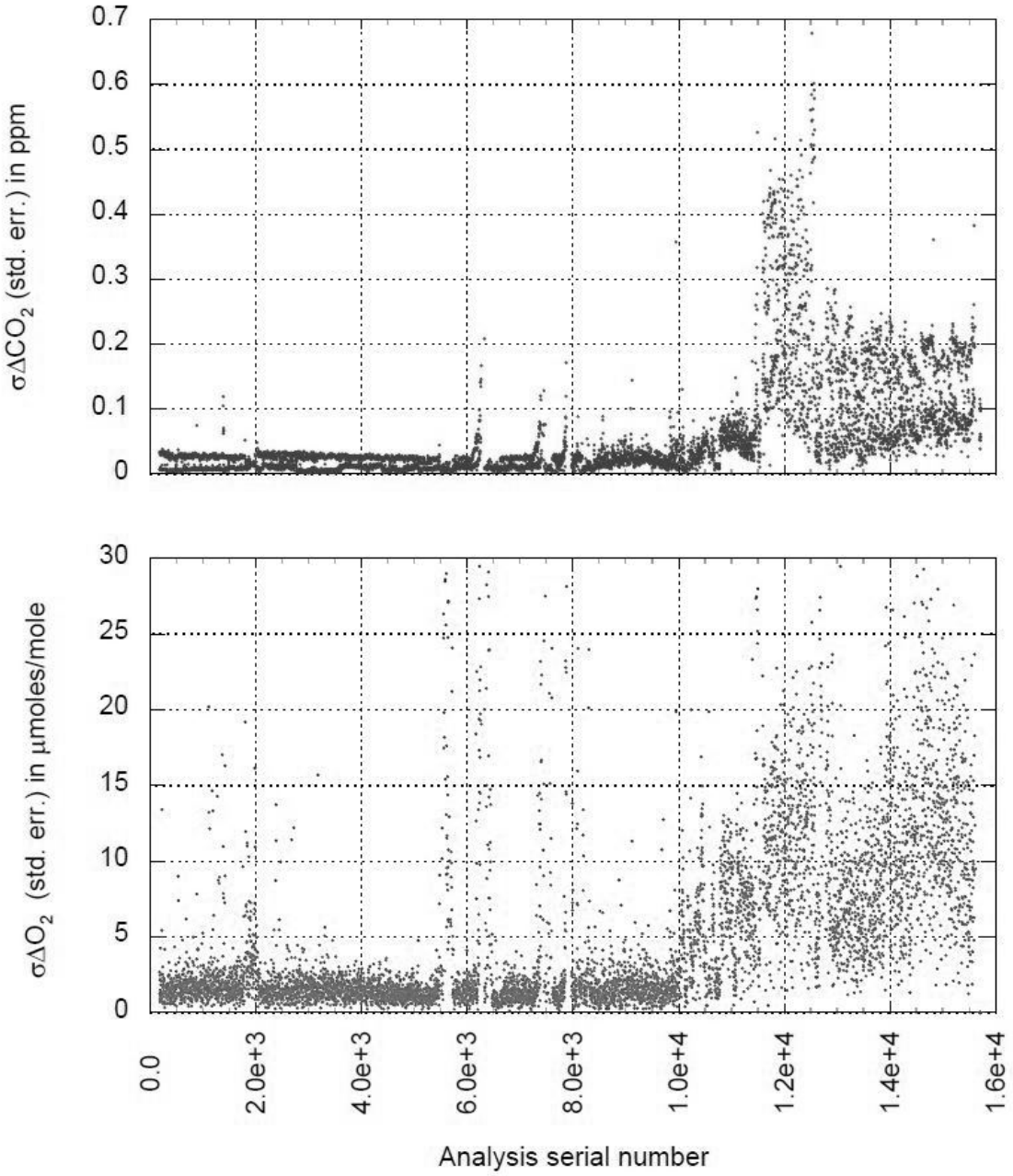

Figure A.4: Error in CO_2 (top) and O_2 (bottom) plotted against serial number, which corresponds to the chronological order of data points. Error varied over time and tended to increase significantly at the end of the study period. We speculate that this increase in error may be due to problems with system gas tightness (Figure 2.5).

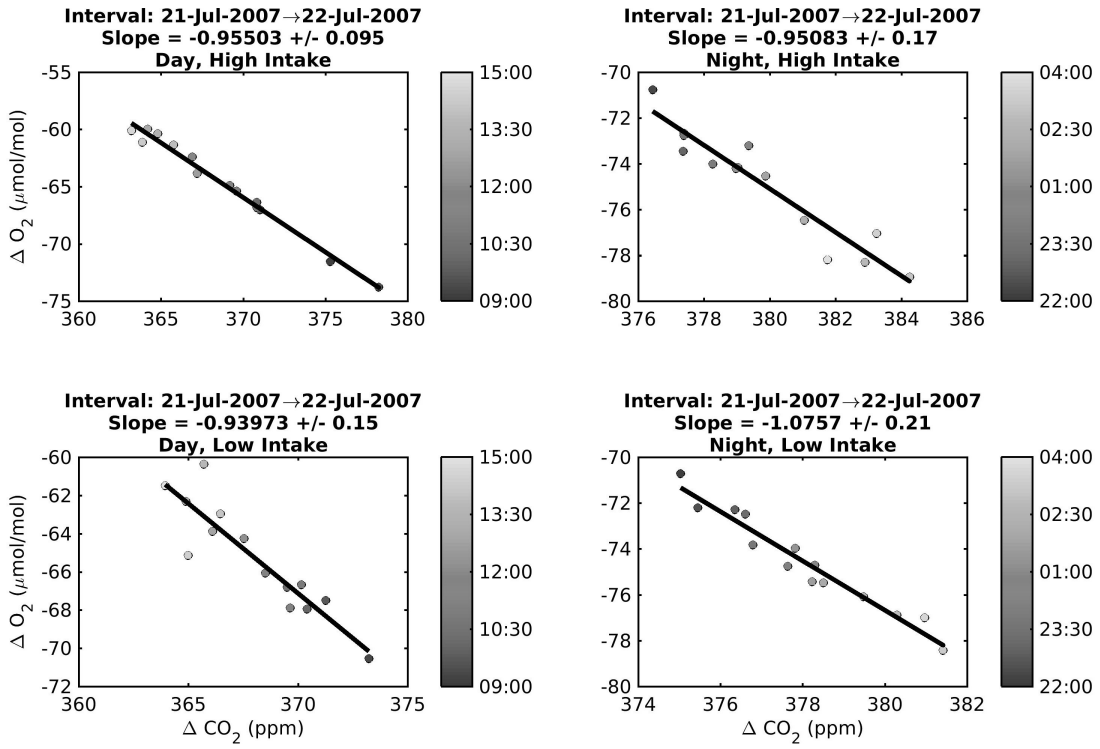

Figure A.5: Example of scatterplots of O_2 vs. CO_2 for the four subsets of data: day and night at high and low intakes. These plots are for 9am–3pm on 21 July and 10pm–4am on 21–22 July in 2007. The shading of the points indicates the time at which each data point was taken (in local time). Note the overall decrease in CO_2 and increase in O_2 during the daytime interval, driven by photosynthesis, and the overall increase in CO_2 and decrease in O_2 during the nighttime interval, driven by respiration. The line shows the results of a Deming regression (described in Section 3.1), and the slope, corresponding to the $O_2 : CO_2$ exchange ratio, is indicated above each plot (Figure 3.1).

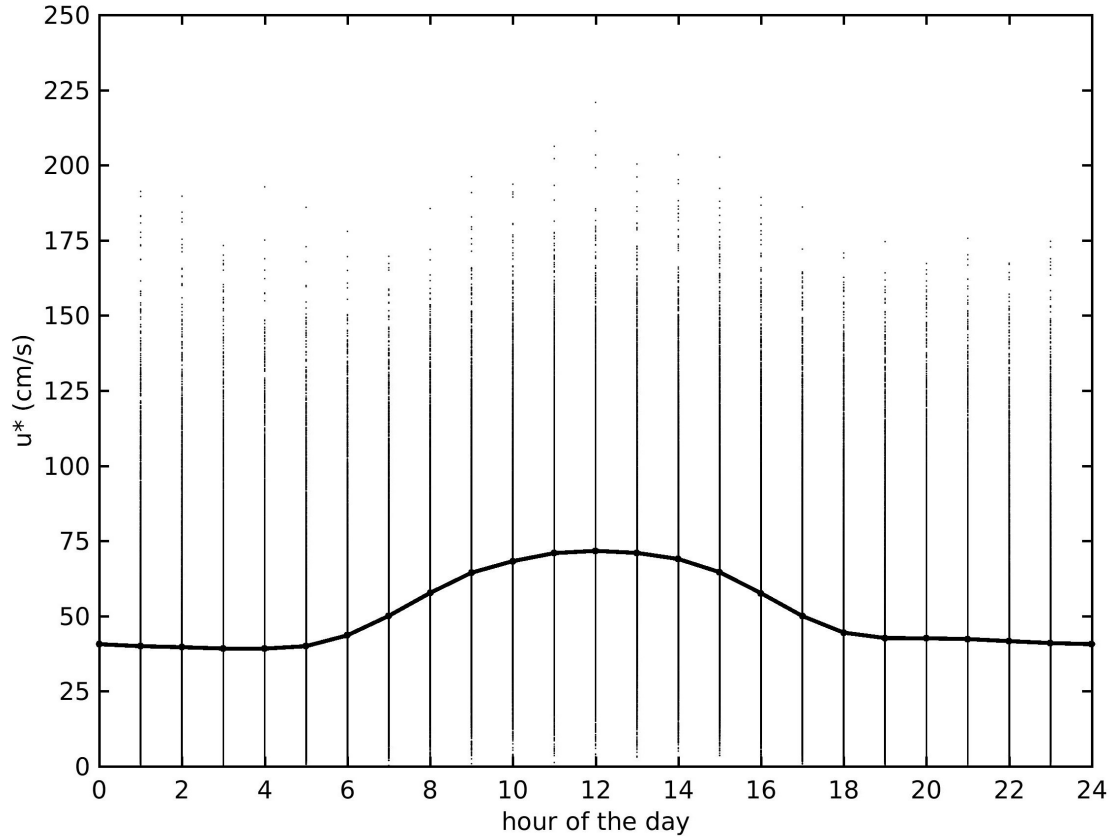

Figure A.6: Hourly values of friction velocity (u*, cm/s). Points show hourly u* values for 1991-2015 measured at the Harvard Forest EMS tower, and the line shows the hourly average. Although u* is highly variable, with a large amount of scatter for each hour of the day, standard errors on the hourly averages are too small to be seen on the plot, and the climatology is well constrained (Figure 3.2).

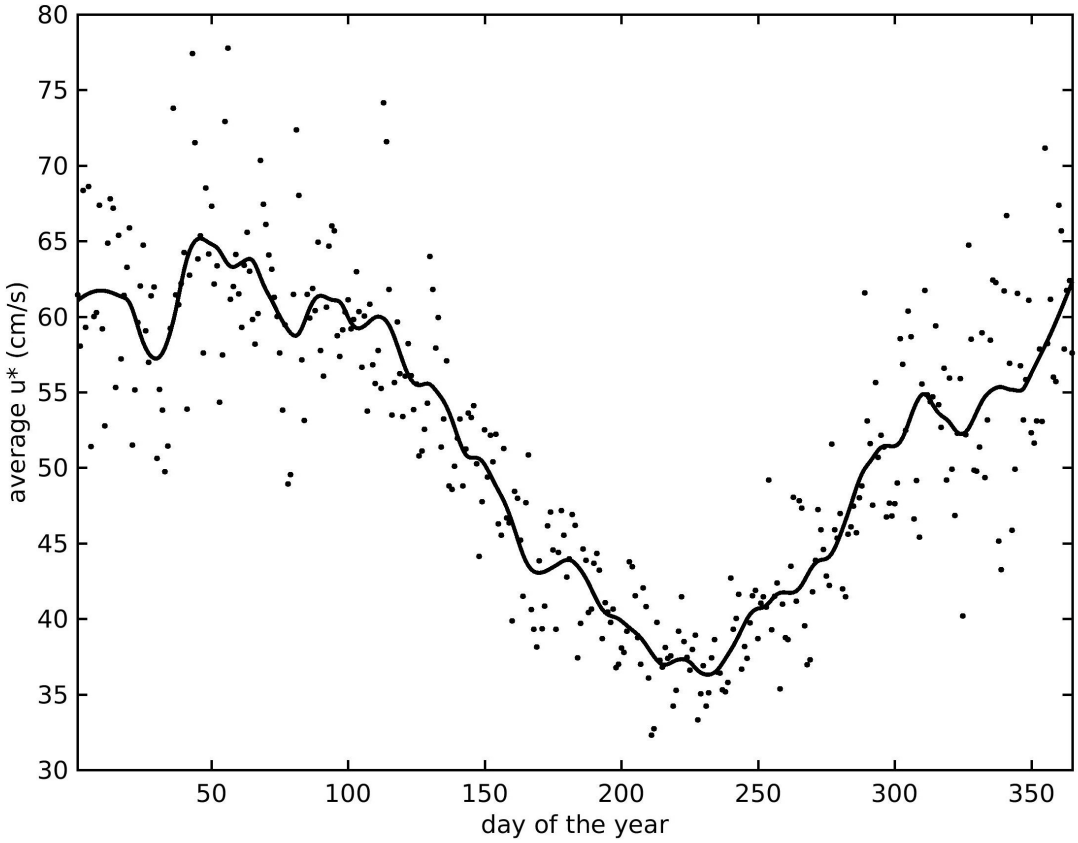

Figure A.7: Seasonal variability in u* (cm/s). Points show the average u* value for each day of the year, calculated from hourly data for 1991-2015. The line shows the results of a loess smoothing function using a window of 36.5 days. Friction velocity tends to be higher and more variable in the winter months compared to the summer (Figure 3.3).

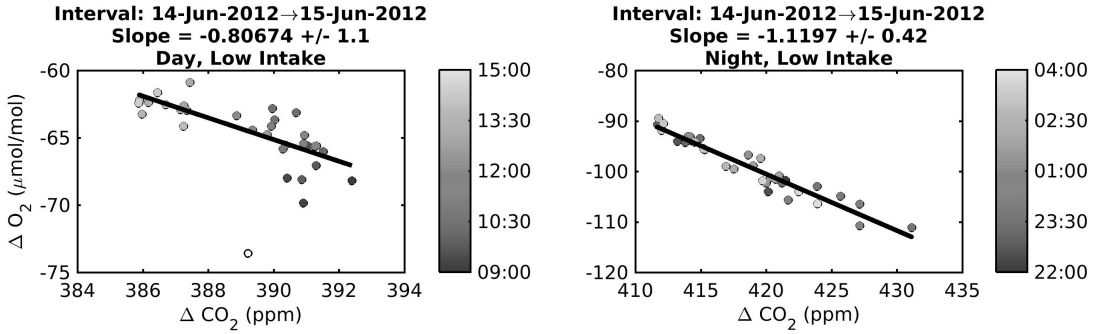

Figure A.8: Scatterplots of O_2 vs. CO_2 for 14–15 June 2012 at the low intake. The empty circle denotes a point that was removed by the orthogonal residual cut and was not included in the final fit shown by the line (Figure 3.4).

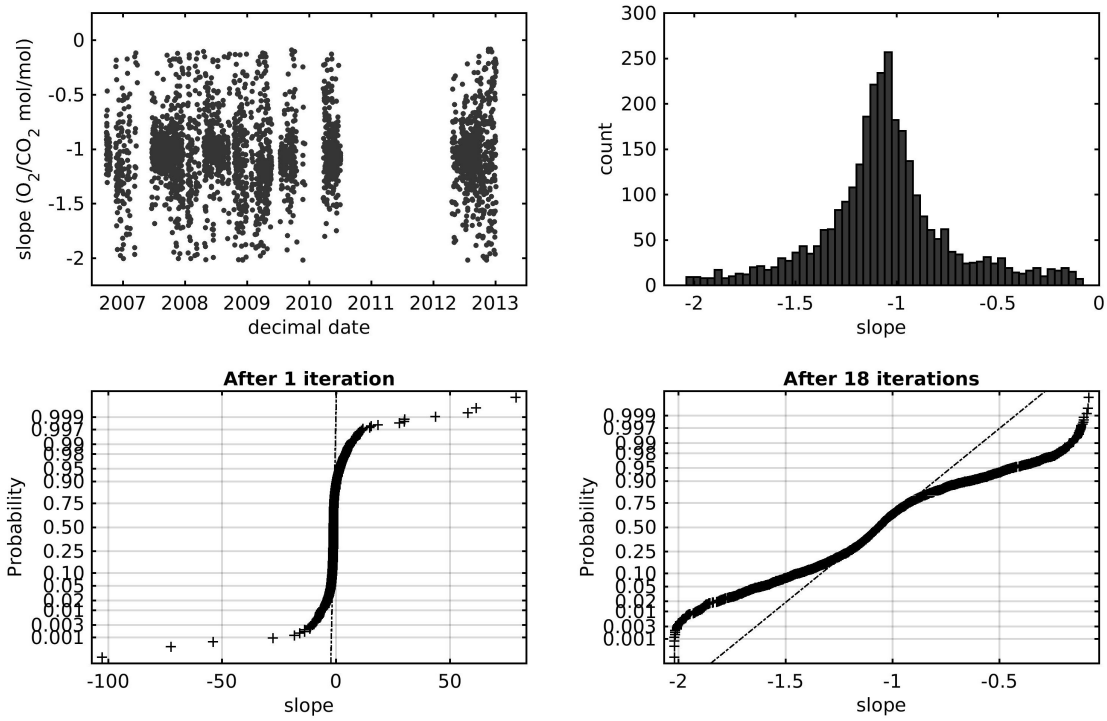

Figure A.9: Distribution of slopes for the entire dataset after the iterative cut. The top left panel shows the iteratively cut slopes plotted against time in decimal years; gaps indicate periods where no data were collected. The top right panel shows a histogram of the iteratively cut slopes. The bottom left panel shows a normal probability plot for the entire dataset after the first iteration of the iterative cut, whereas the bottom right shows the same plot after the iterative cut was completed, demonstrating the reduction of the dataset's non-Gaussian tails. The dotted lines in the normal probability plots indicate the point pattern expected for a normal distribution. Summary statistics are in Table 3.1 (Figure 3.5).

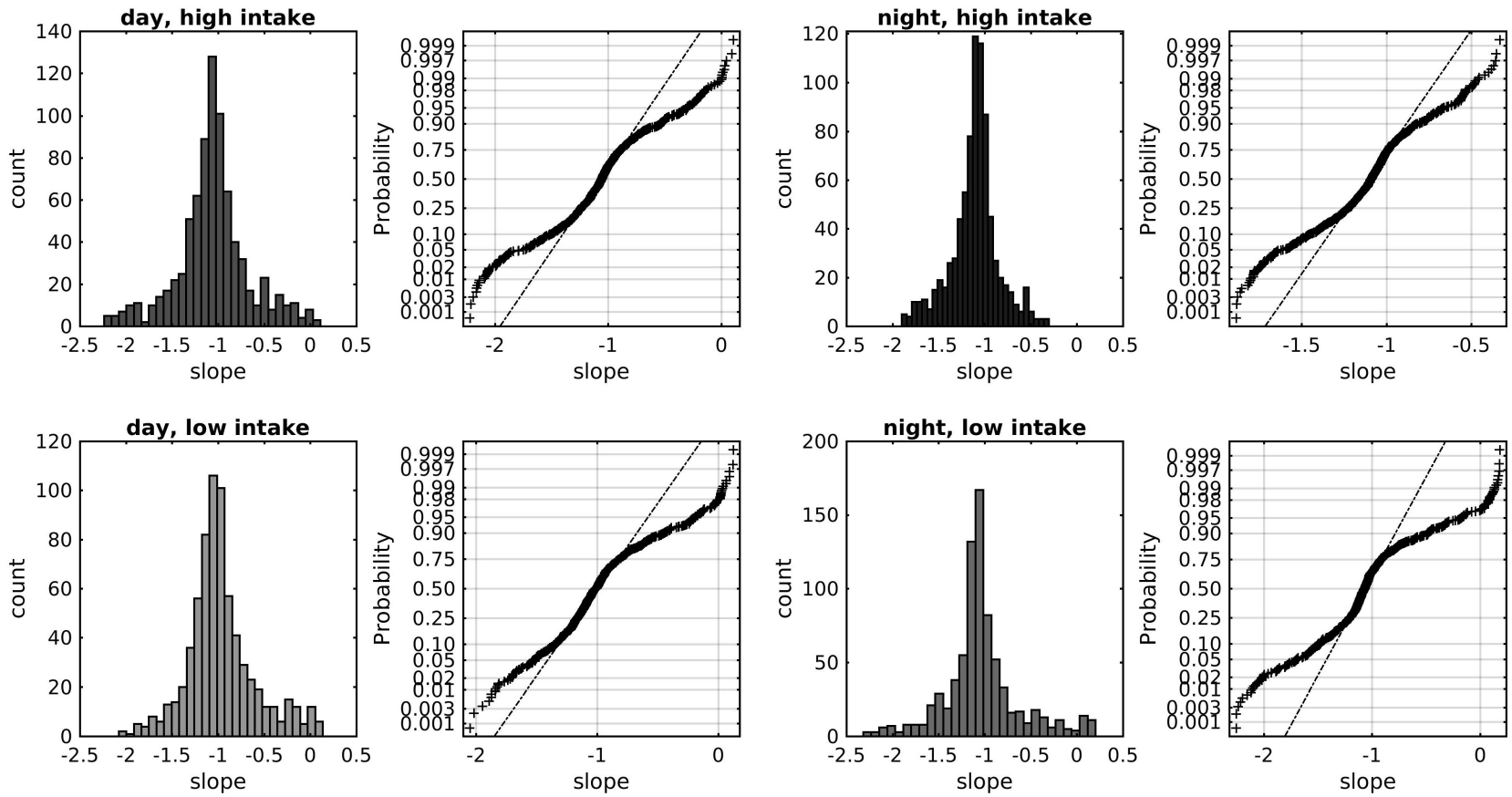

Figure A.10: Histograms and normal probability plots for the four data subsets. Each normal probability plot corresponds to the histogram to its left. Although the iterative cut partially normalizes the distributions, there are still non-Gaussian tails, visible in the normal probability plots, in which the dotted line indicates the point pattern expected for a normal distribution. For this reason, we include both the means and the medians for comparison in the summary statistics in Table 3.1 (Figure 3.6).

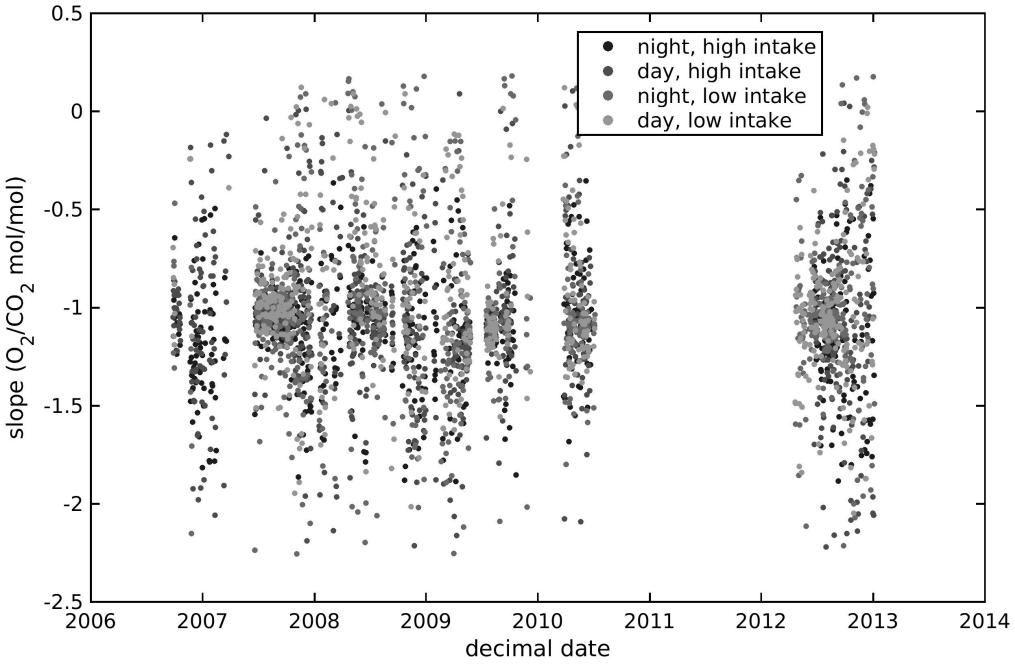

Figure A.11: Plot of slopes vs. time in decimal date. Point shading indicates the time interval and intake height for each data point. The increase in spread towards the end of the study period may be a reflection of the increase in measurement error shown in Figure 2.5 (Figure 3.7).

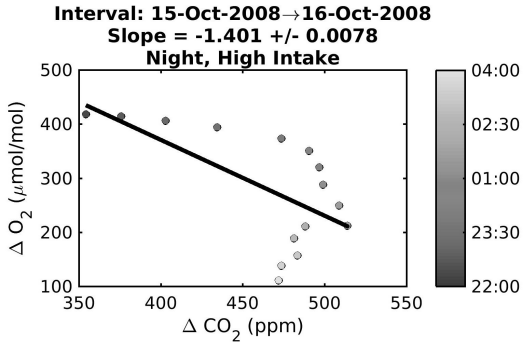

Figure A.12: Scatterplot of O_2 vs. CO_2 for the nighttime interval on 15–16 October 2008. Although the linear fit is obviously inappropriate for this plot, the uncertainty in the regression is low. This is due to the anomalously large ranges in CO_2 and O_2, which are much higher than the ranges characteristic of the majority of the plots (Figure 3.8).

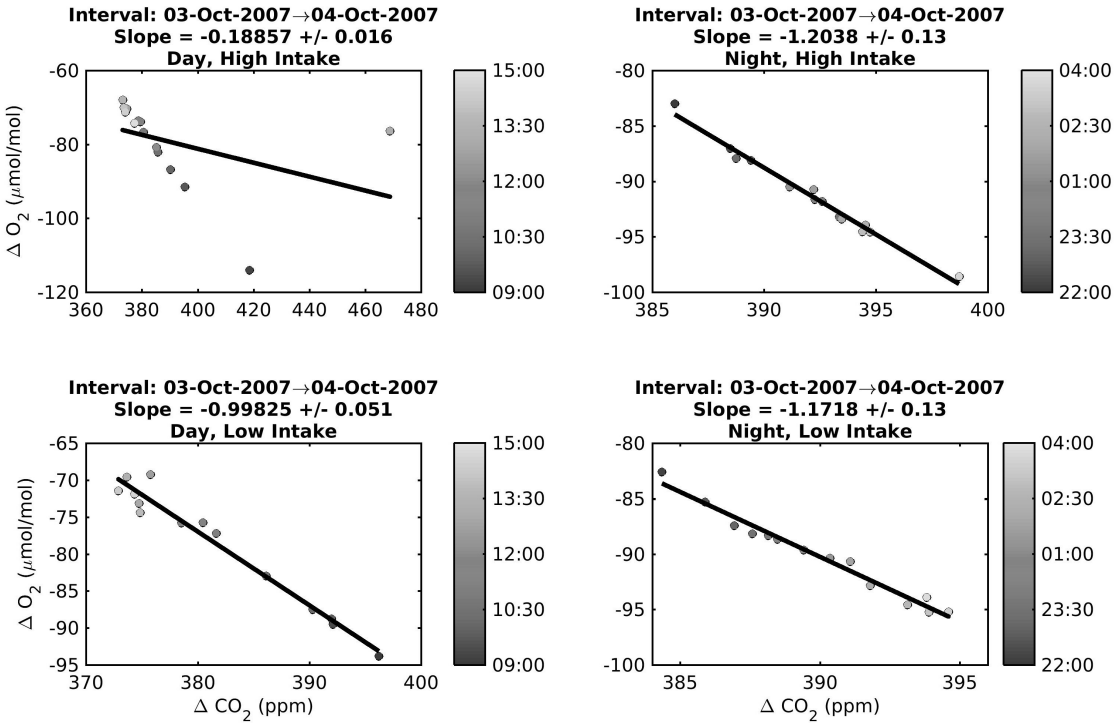

Figure A.13: Scatterplots of O_2 vs. CO_2 for 3–4 October 2007. In the "Day, High Intake" plot, an anomalously high CO_2 value resulted in a non-representative fit without much increasing the uncertainty, while the other subsets, whose linear fits are much more representative of the trend in the data, have higher uncertainties. Similar to Figure 3.8, a point with a much higher CO_2 or O_2 value than the rest of the data causes problems for the linear regression and results in low uncertainties that do not reflect the actual quality of the fit (Figure 3.9).

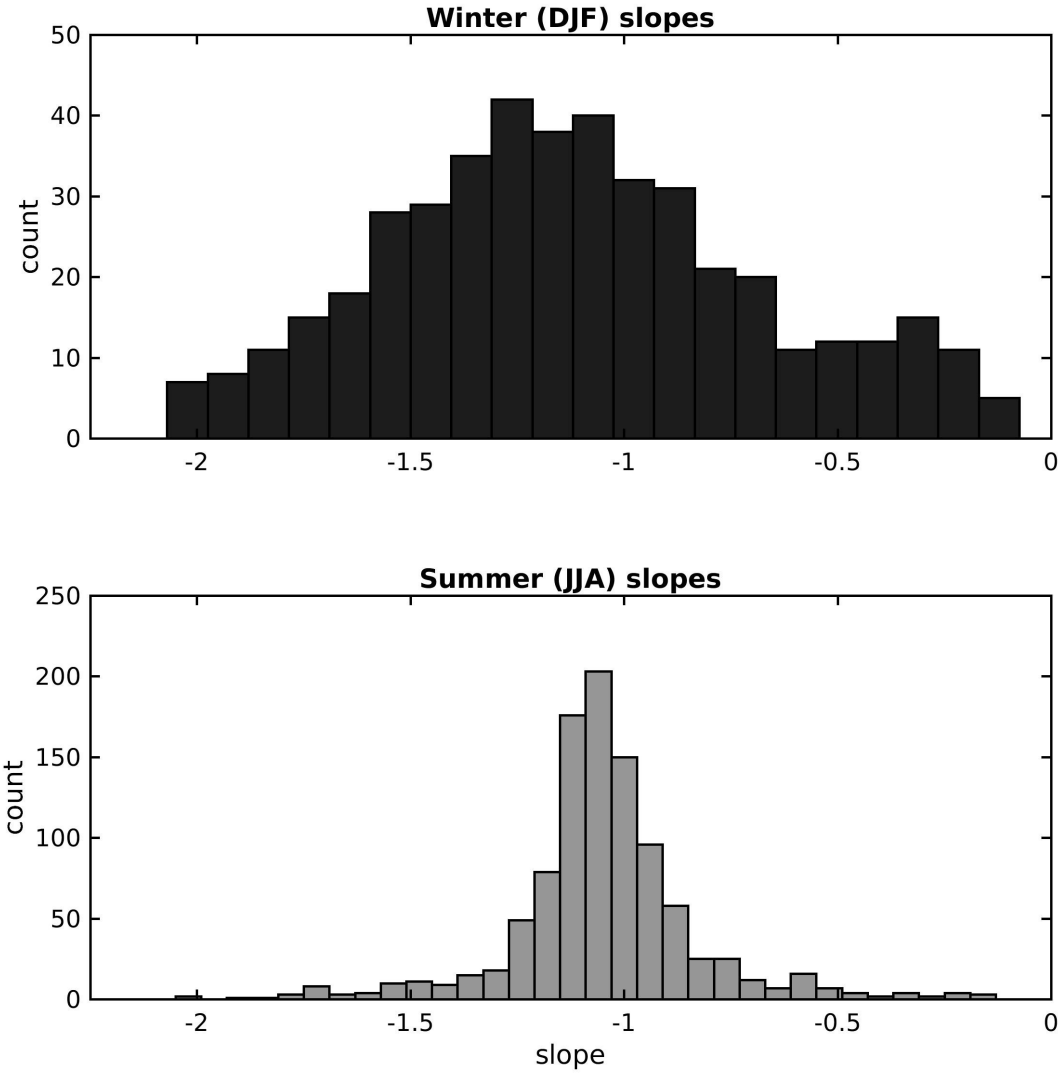

Figure A.14: Histograms for winter and summer slope values. Winter comprises the months of December, January, and February and summer comprises June, July, and August. Winter slopes have a more negative mean value, closer to the stoichiometric ratio for fossil fuel combustion, although the sample size is smaller for winter than for summer. Statistics are shown in Table 4.1 (Figure 4.4).

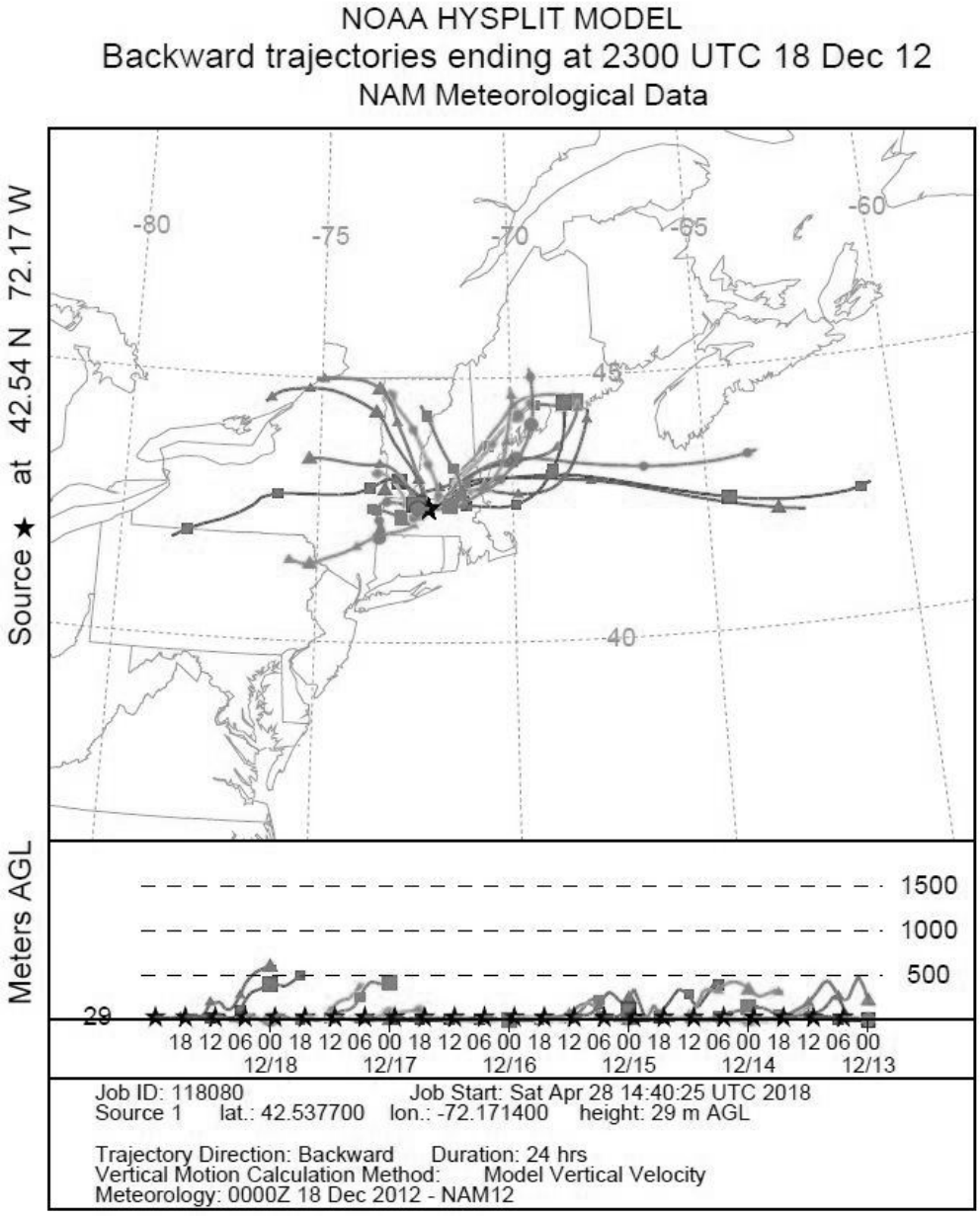

Figure A.15: 24-hour back-trajectories for Harvard Forest produced by the HYSPLIT model, starting on 18 December 2012. Each trajectory shows the path of an air parcel during the 24 hours before it arrived at Harvard Forest. Trajectories were started at 6-hour intervals, moving back in time, to create the ensemble of trajectories shown. For this longer time period, the source region is highly variable and encompasses a large area, demonstrating our rationale for using a shorter 6-hour interval for our slope plots to reduce source region variability (Figure 5.1).

Figure A.16: A set of 48 6-hour back-trajectories produced by the Hysplit model, plotted on satellite imagery from Google Earth. Each trajectory shows the path of an air parcel during the 6-hour period before it arrives at Harvard Forest. Trajectories include 24 daytime and 24 nighttime trajectories for two days in winter and two days in summer. For each day, trajectories were started each hour between 9am–3pm and 10pm–4am, the same time intervals used for the slope plots. These 6-hour intervals encompass a fairly consistent source region with a radius of 50–200 km. The area is mostly forested with some small urban centers (Figure 5.2).

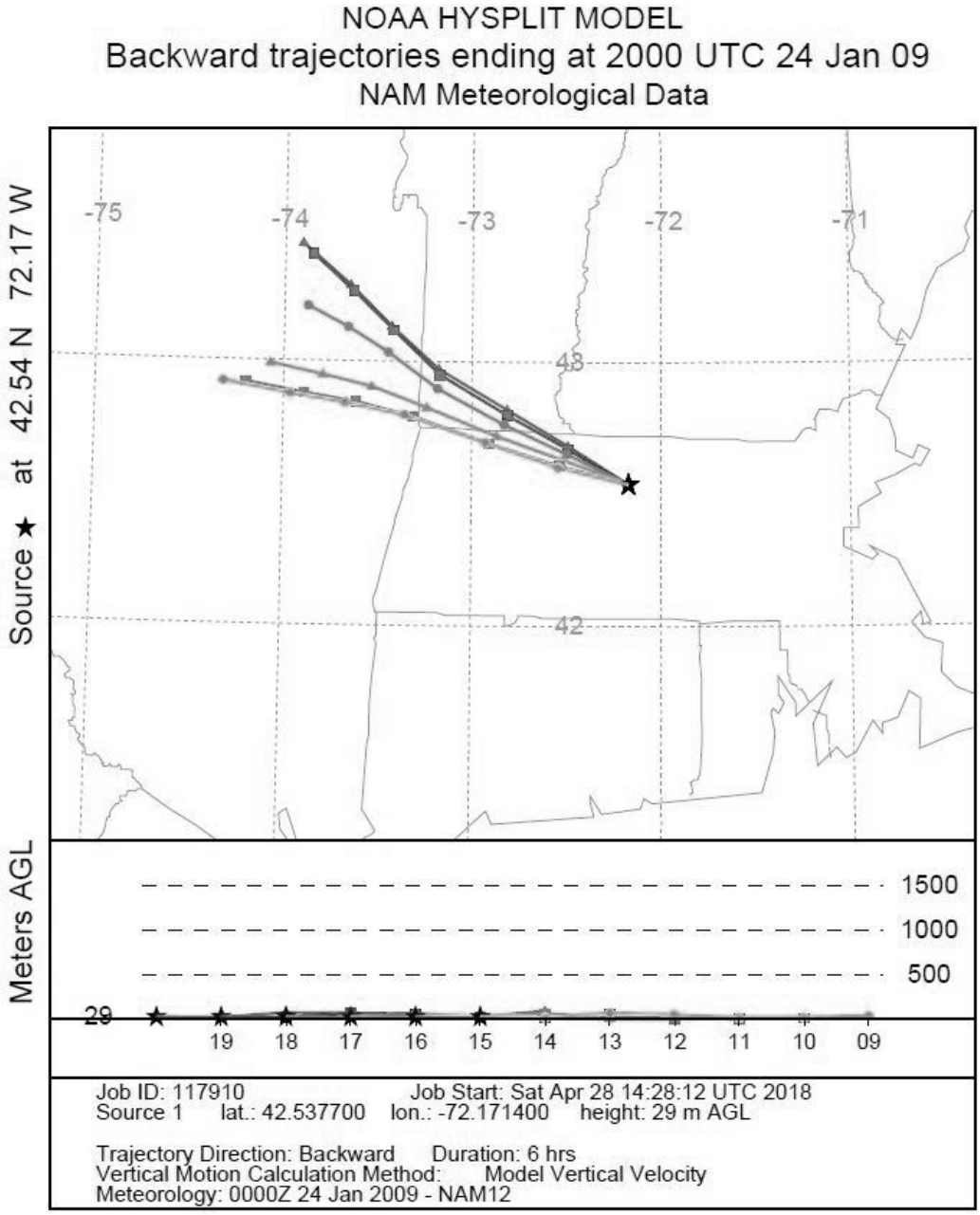

Figure A.17: 6-hour back trajectories for Harvard Forest on 24 Jan 2009, with one trajectory started each hour between 9am and 3pm EST (corresponding to the 6-hour daytime interval used to calculate slopes), showing variability in the source region over that time period. This particular day shows a steady shift in the source region over time, although the path length and general source area remain similar. For this period, background variability might have a slightly stronger influence on the observed CO_2 and O_2 variability at Harvard Forest (Figure 5.3).

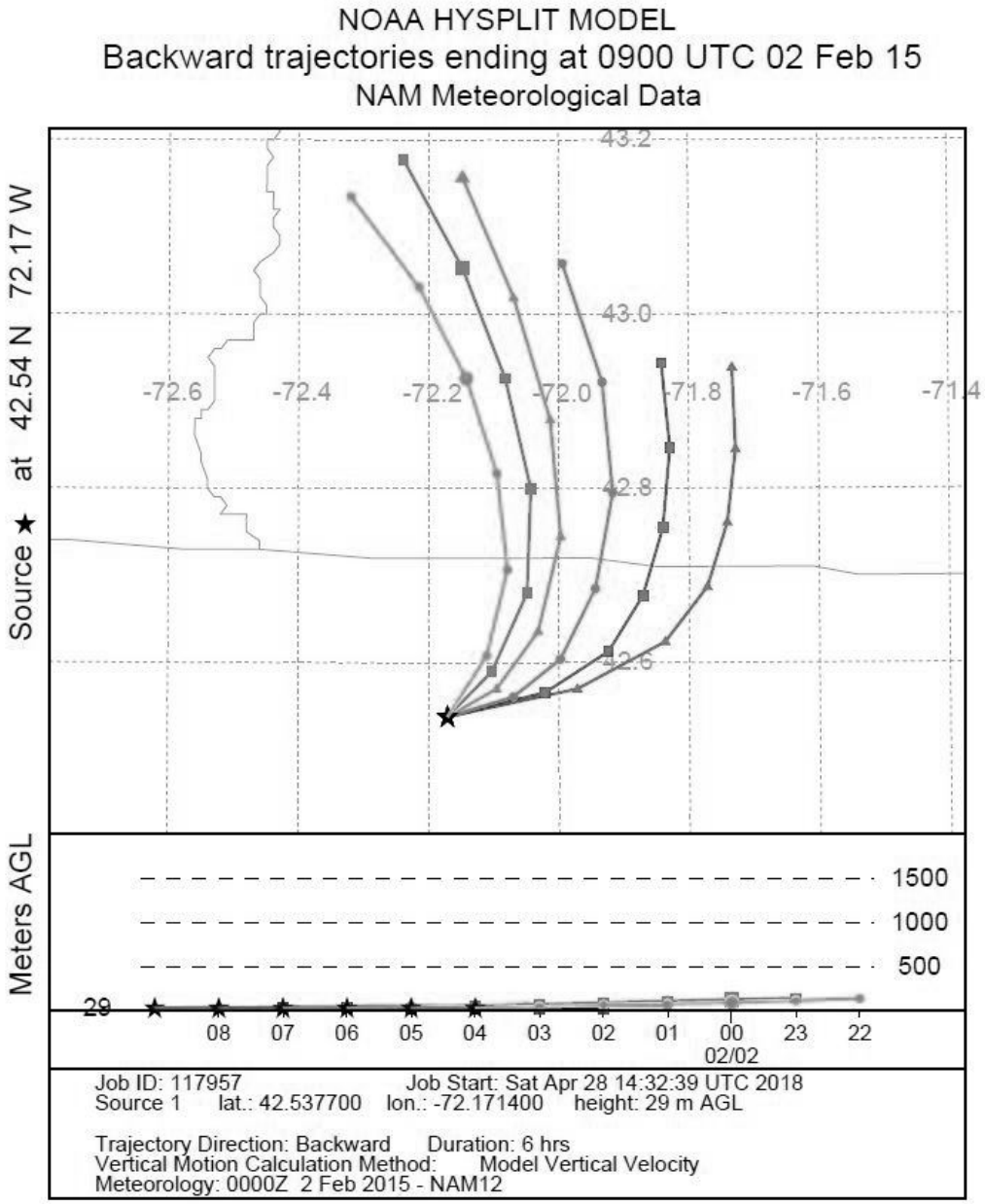

Figure A.18: 6-hour back trajectories for Harvard Forest on 2 Feb 2015, with one trajectory started each hour between 10pm and 4am EST (corresponding to the 6-hour nighttime interval used to calculate slopes), showing variability in the source region over that time period. For this particular night, changes in the source region were minimal, with air coming from a small area in southwestern NH, so local influence would likely dominate the observed CO_2 and O_2 variability at Harvard Forest (Figure 5.4).

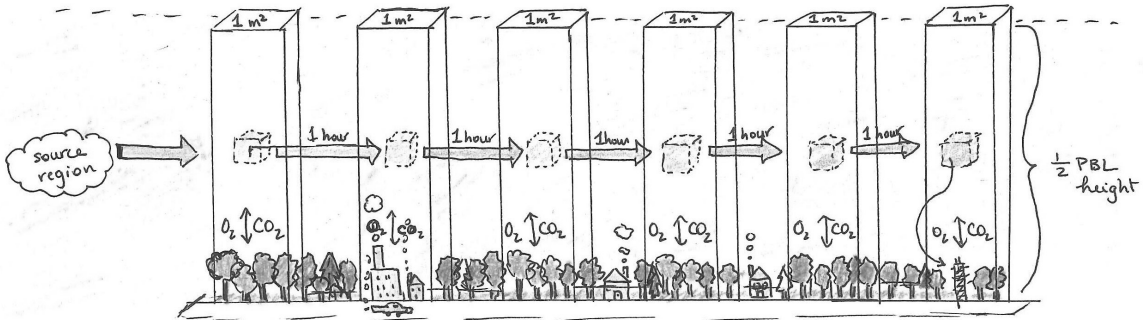

Figure A.19: A conceptual diagram of the box model. Only 6 of the 30 boxes are shown, denoting the locations of the air parcel at each hour along its path to Harvard Forest. Air enters the model from the left (the source region) and leaves the model on the right (at Harvard Forest). As the air parcels travel through each box, O_2 and CO_2 are exchanged between the boxes and the landscape, which changes the mixing ratio of the parcel. The parcels shift forward every 12 minutes, and their mixing ratios are recorded as they arrive in the final "Harvard Forest" box on the right (Figure 5.5).

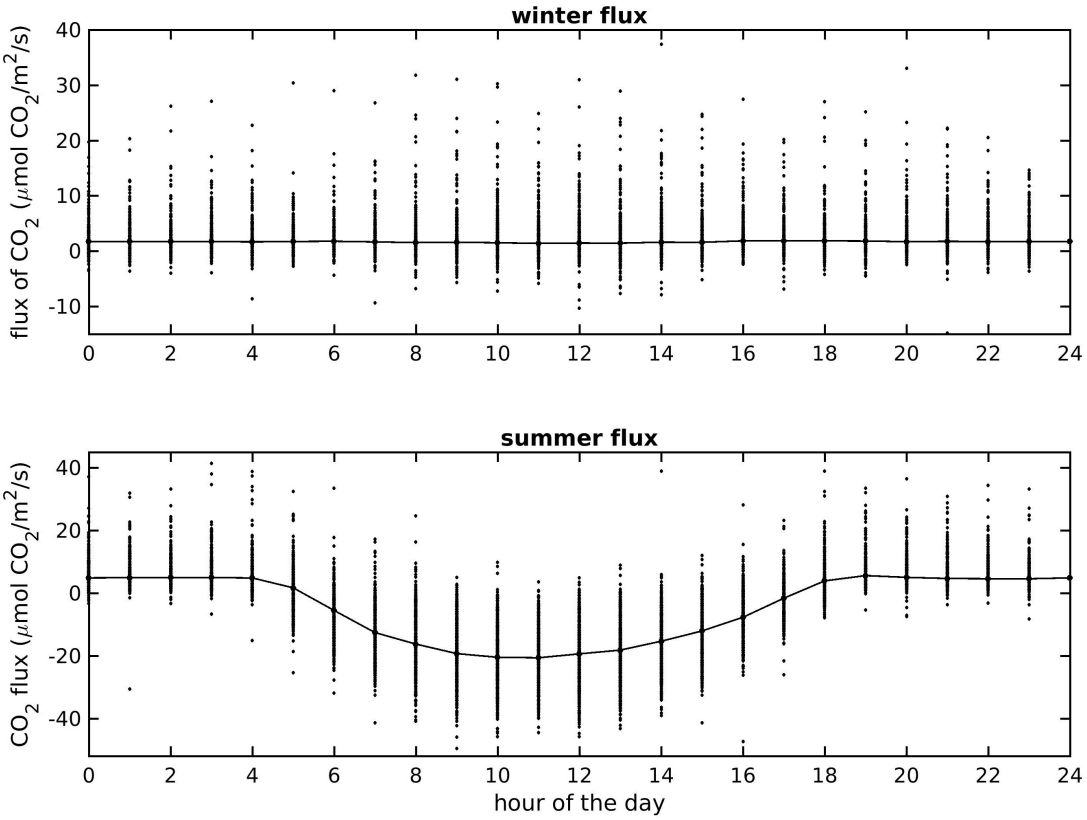

Figure A.20: An hourly climatology of CO_2 fluxes at Harvard Forest based on hourly eddy flux data from 1991–2015, separated into winter and summer climatologies. Points show the ensemble of hourly data points, and the line shows the average value for each hour. Summer fluxes have much higher magnitudes and a more pronounced daily cycle than winter fluxes, which are low, slightly positive (due to respiration) and fairly constant. One outlier in the summer plot (hour 23, flux 107.7 μmol $CO_2/m^2/s$) was omitted for clarity (Figure 5.6).

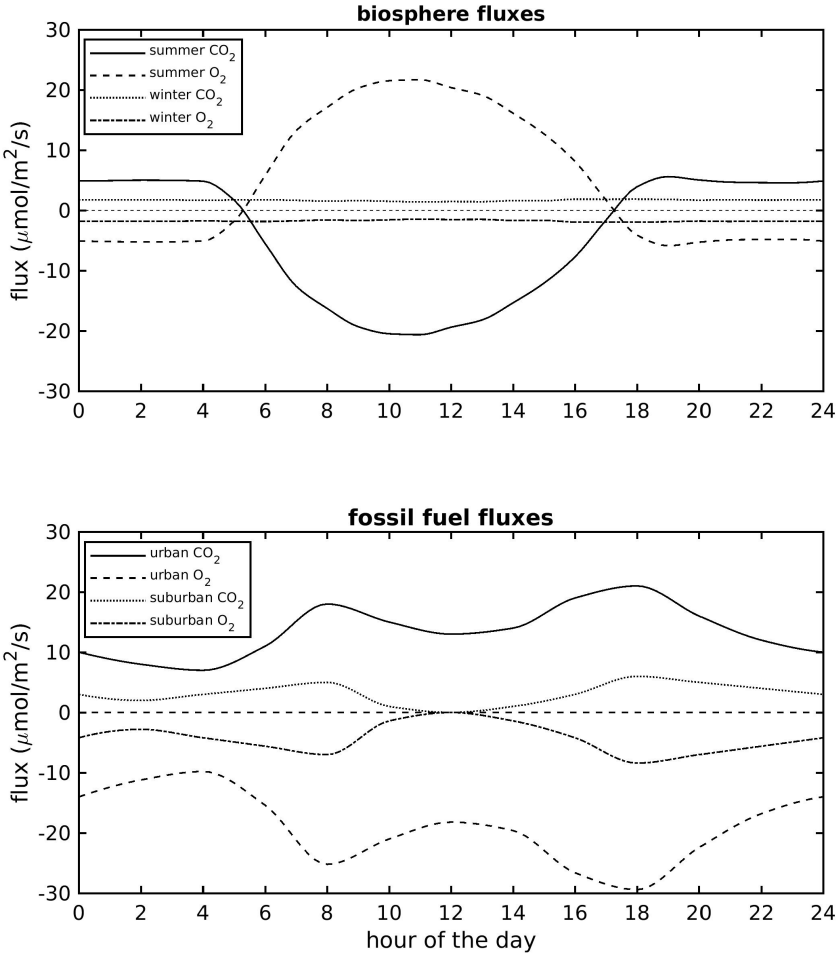

Figure A.21: CO_2 and O_2 flux cycles used as input for the box model. Lines show CO_2 fluxes and corresponding O_2 fluxes calculated using the appropriate stoichiometric ratio. The top panel shows forest fluxes, separated into winter and summer cycles, and the bottom panel shows fossil fuel fluxes for an urban and suburban area. Although urban and suburban cycles do vary seasonally, the change is not as pronounced as the biospheric variability and is not essential for our simple model (Figure 5.7).

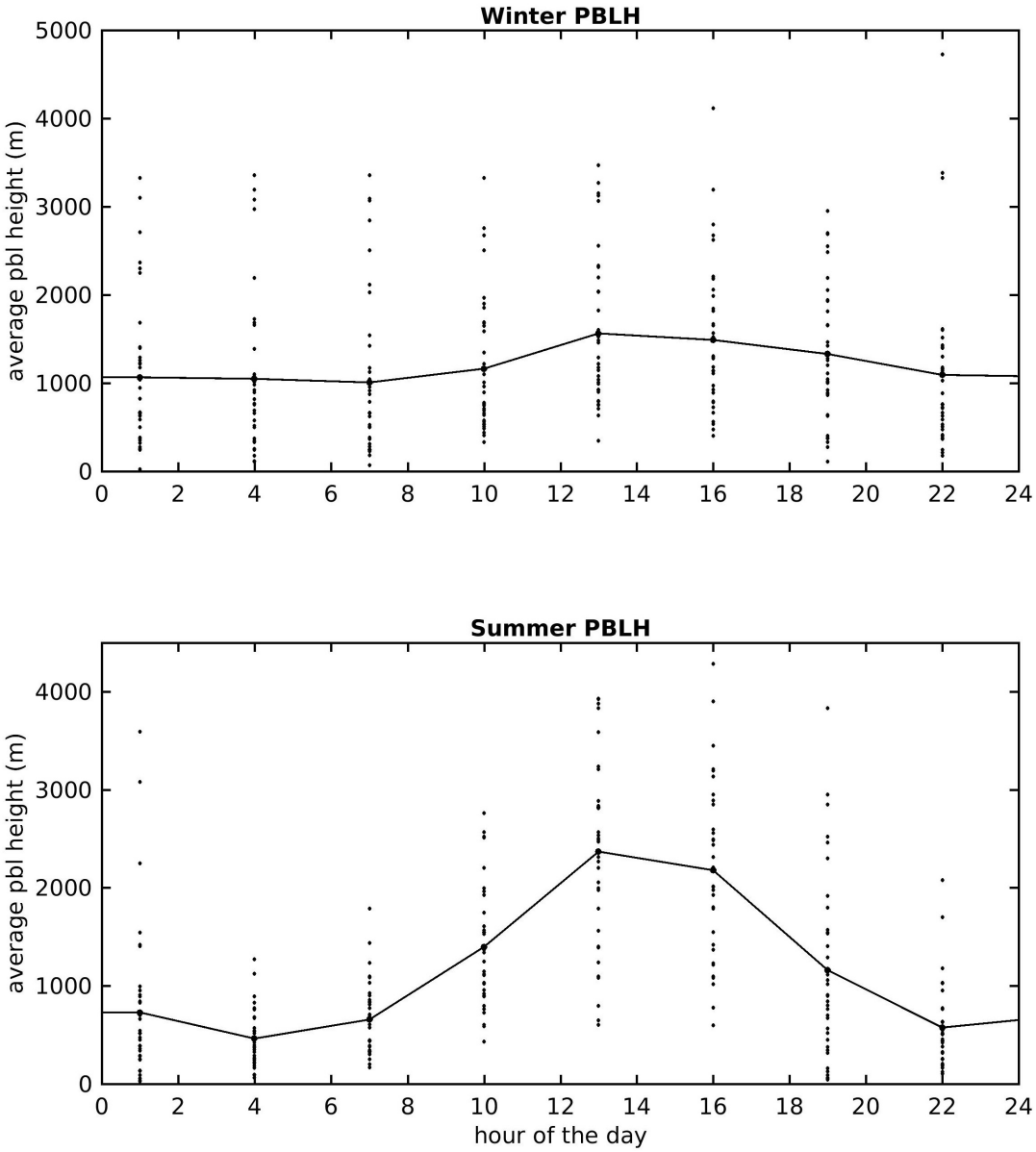

Figure A.22: Daily cycles in PBL height for summer and winter calculated from ARL data [*NOAA*, 2017]. The points show hourly values (with a resolution of three hours) for 36 days during each season, and the line shows the average for each hour (with points every three hours connected by a linear interpolation). The summer PBL height is more variable, with lows in the morning and highs in the afternoon driven by radiative heating of the land, whereas the winter profile is relatively flat (Figure 5.8).

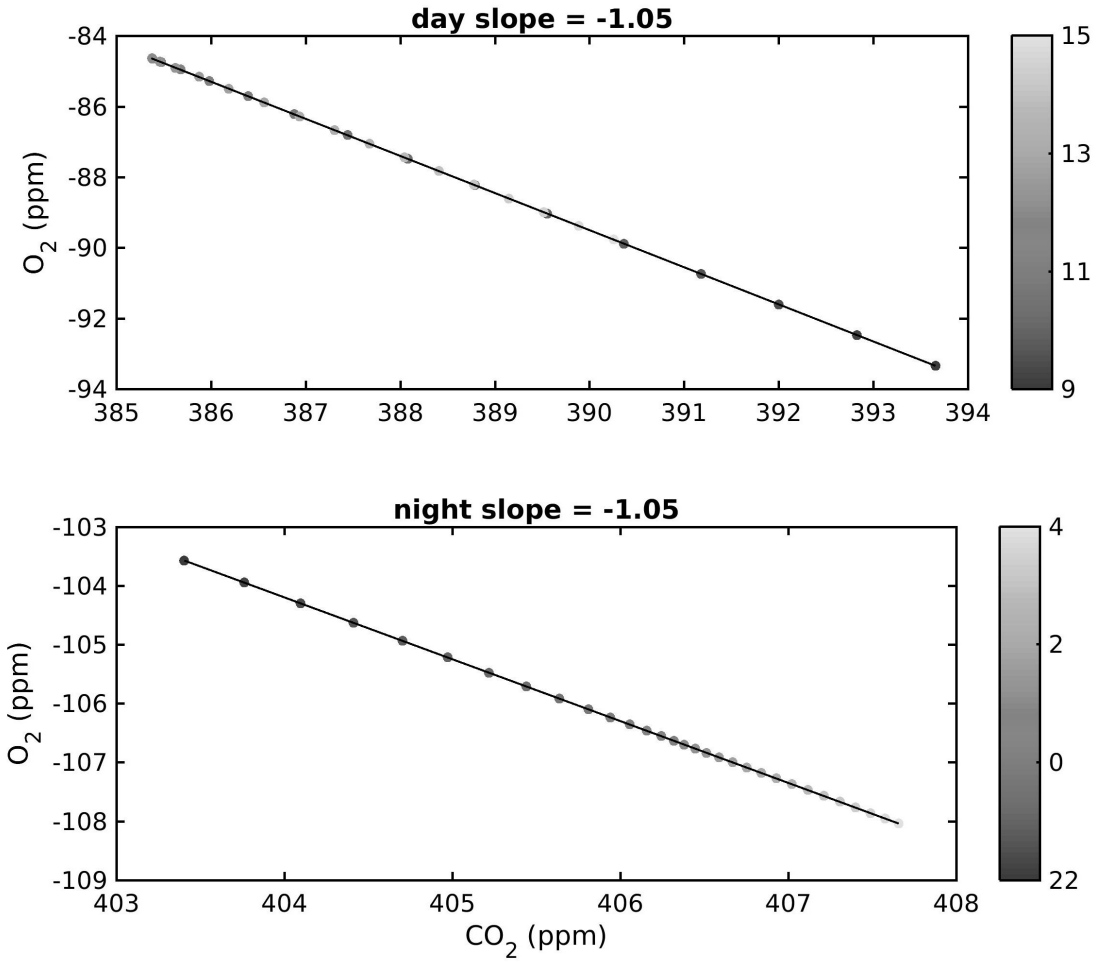

Figure A.23: Model output for a forested landscape, showing the resulting scatterplots of O_2 vs. CO_2 for the same daytime and nighttime periods used with the observed data. The shading of the points indicates the time each arrived at Harvard Forest, with the labels on the colorbar denoting hours of the day. The line shows the linear regression associated with the slope value. This plot was produced using the summer PBL height and summer forest flux cycle. A completely forested landscape produces the expected slope of -1.05, reflecting the stoichiometric ratio of the forest fluxes used to drive the model (Figure 5.9).

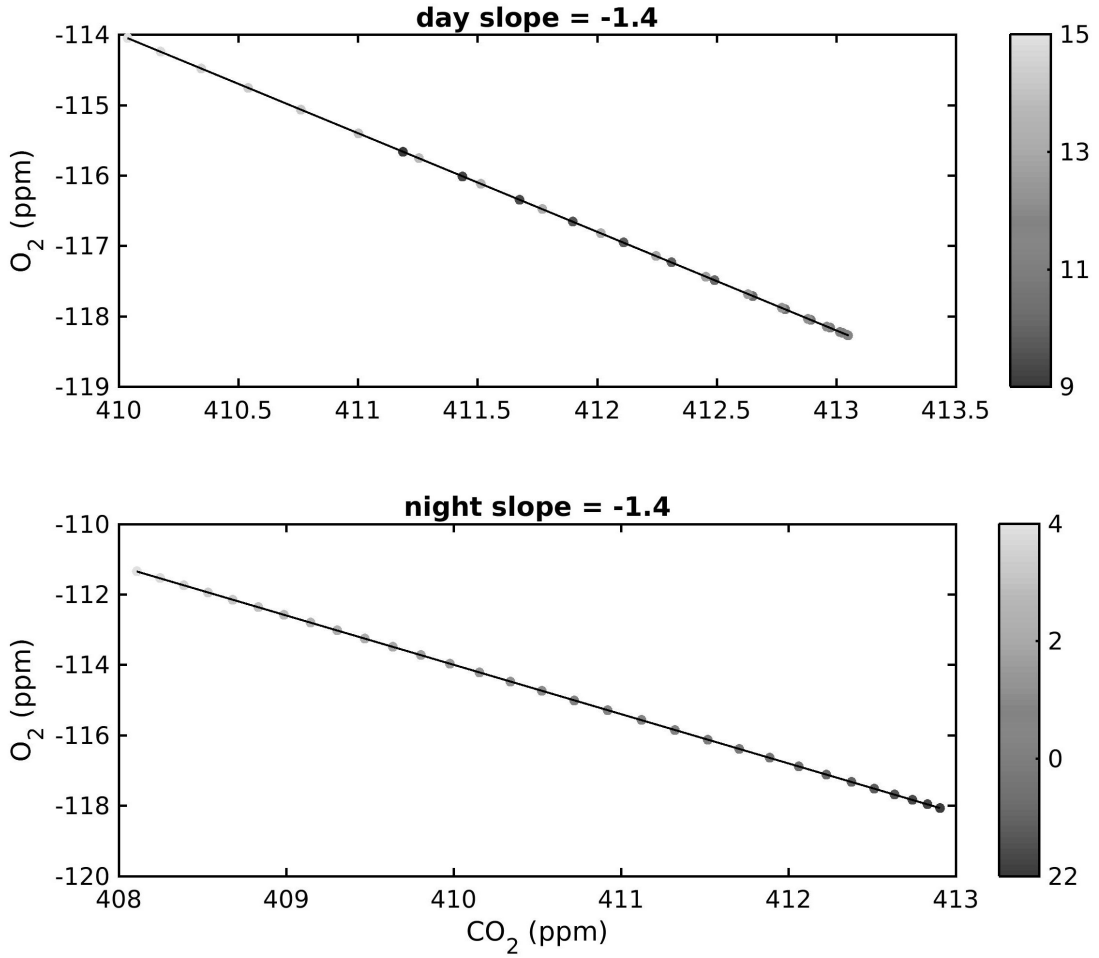

Figure A.24: Model output for an urban landscape, using the winter PBL height and urban flux cycle. A completely urban landscape produces the expected slope of -1.4 (Figure 5.10).

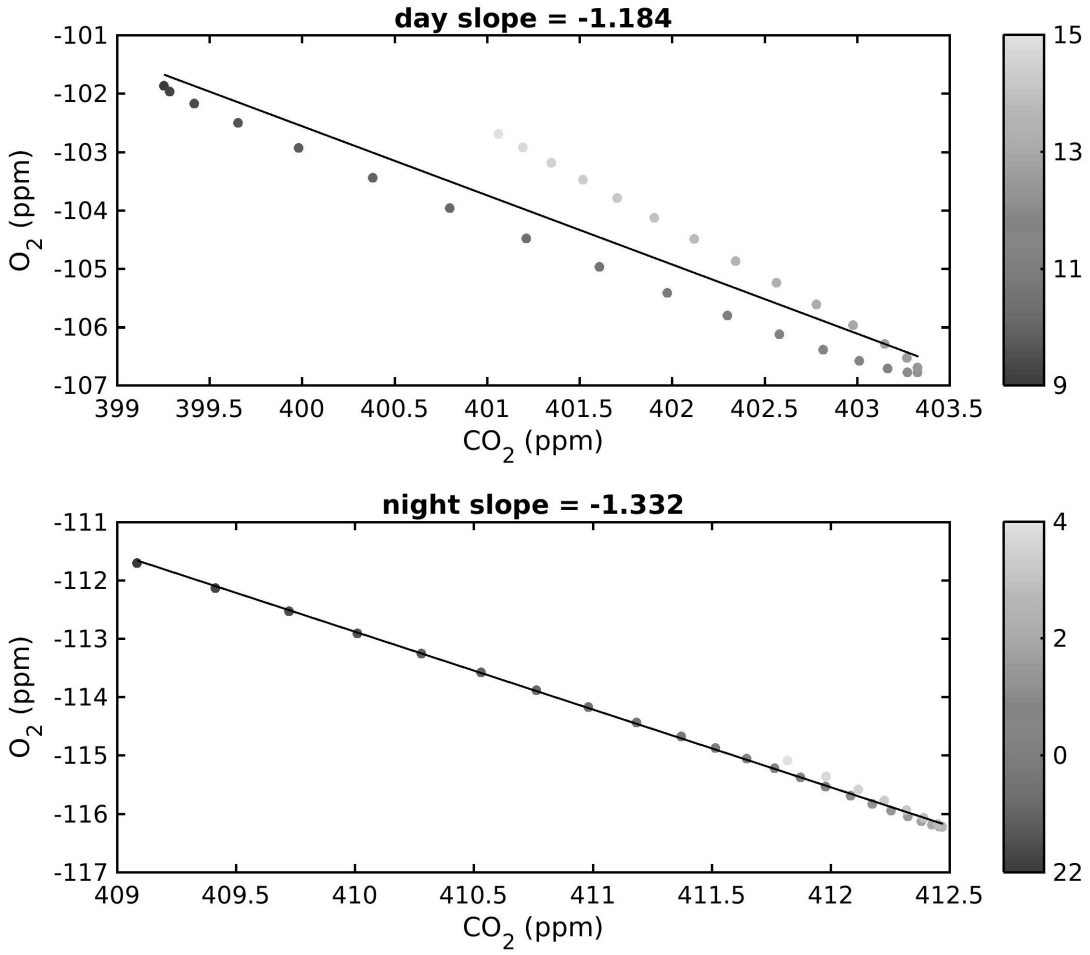

Figure A.25: Model output for a mixed landscape consisting of 15 boxes of urban landscape followed by 15 boxes of summer forest, using the summer PBL height cycle. Because the summer forest fluxes are of similar magnitude to the urban fluxes, the slope produced is intermediate between -1.05 and -1.4, the input stoichiometries of the two landscape types (Figure 5.11).

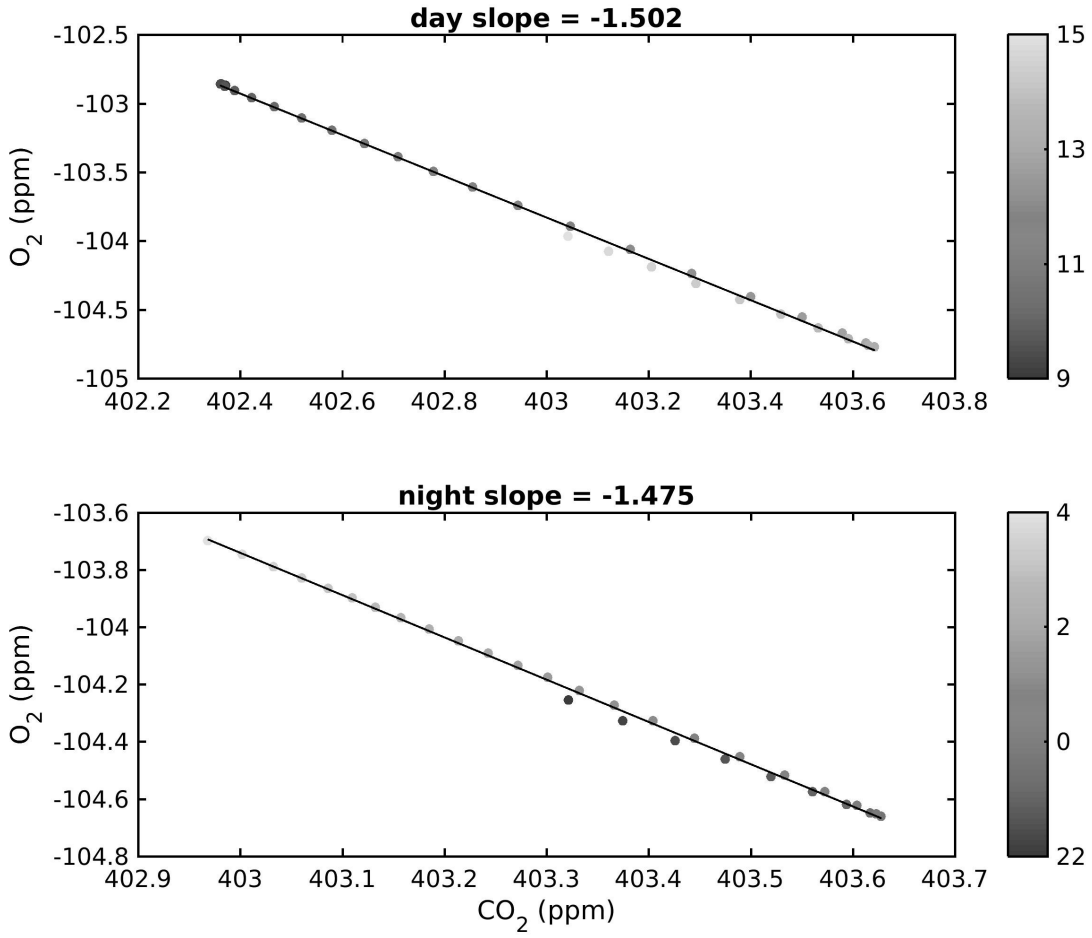

Figure A.26: Model output for a mixed landscape consisting of 5 boxes of urban landscape followed by 25 boxes of winter forest, using the winter PBL height cycle. The resulting slopes are outside the range of the two input stoichiometries (-1.05 to -1.4), because the use of a constant source term for the model results in alternating increases and decreases in CO_2 and O_2 throughout the day based on the daily flux cycles, producing point patterns outside of a straight line and slope values outside the range of the inputs (Figure 5.12).

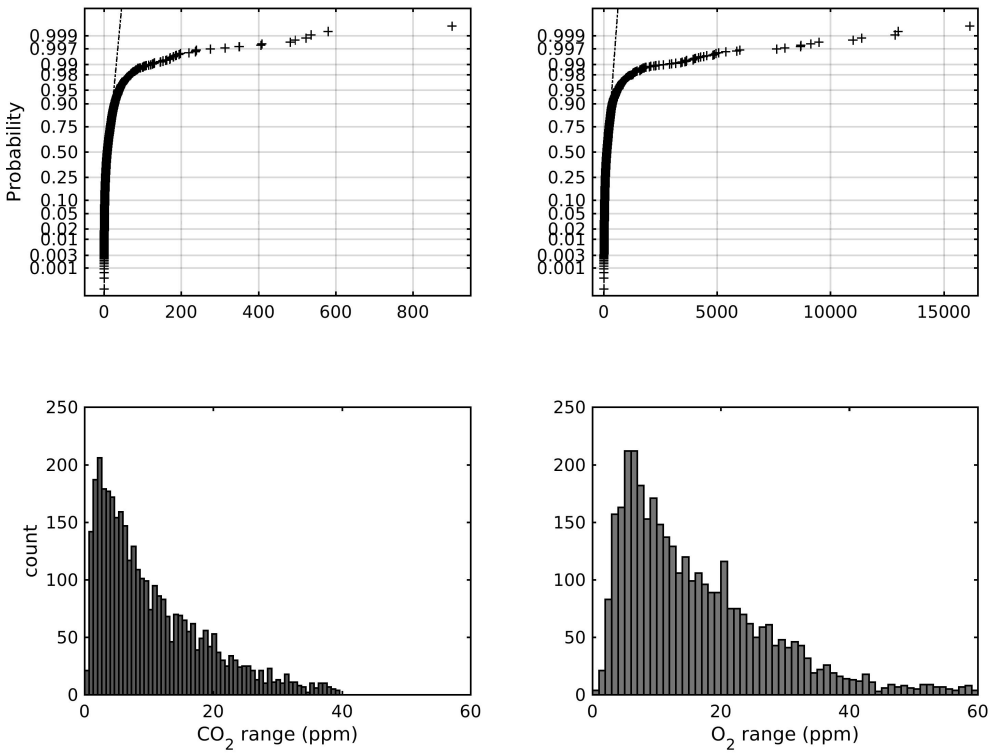

Figure A.27: Histograms showing the distribution of CO_2 and O_2 ranges for the full set of 6-hour slope values. Top two panels show normal probability plots, where the dashed line describes a normal distribution. CO_2 ranges are slightly lower than O_2 ranges overall. One outlier in the normal probability plots with a CO_2 range of 15080 ppm and an O_2 range of 24760 μmol/mol was omitted for clarity, and the upper tail of the histograms was omitted to better show the distribution of the core of the data (Figure 5.13).

Acknowledgements

I would like to acknowledge Mark Battle for his invaluable support, encouragement, and assistance with my honors project as well as throughout my time at Bowdoin. I would also like to thank Bill Munger, Maryann Sargent, Ken Davis, and Susan Solomon for helpful input and ideas about the project. Thanks to Taylor Jones for providing us with code to access ARL archive data. Thanks also to DJ Merrill for help with the database and computer-related questions. This project relied on the previous work of many students, including John Carpenter, Becca Perry, Eric Sofen, Sasha David, Zane Davis, Ryan Hart, Jayme Woogerd, Jakob Scheckman, Sam Seekins, and Karina Graeter.

References

Battle, M., M. L. Bender, P. P. Tans, J. W. C. White, J. T. Ellis, T. Conway, and R. J. Francey, Global carbon sinks and their variability inferred from atmospheric O_2 and $\delta^{13}C$, *Science*, *287*, 2467–2470, 2000.

Bergeron, O., and I. B. Strachan, CO_2 sources and sinks in urban and suburban areas of a northern mid-latitude city, *Atmospheric Environment*, *45*, 1564–1573, 2011.

Blaine, T. W., R. F. Keeling, and W. J. Paplawsky, An improved inlet for precisely measuring the atmospheric Ar/N_2 ratio, *Atmospheric Chemistry and Physics*, *6*, 1181–1184, 2006.

de Araùjo, A. C., K. Bart, A. D. Nobre, A. J. Dolman, M. J. Waterloo, E. J. Moors, and J. S. de Souza, Nocturnal accumulation of CO_2 underneath a tropical forest canopy along a topographical gradient, *Ecological Applications*, *18*, 1406–1419, 2008.

Deming, W. E., *Statistical adjustment of data*, Dover Publications, New York, 2011.

Gallagher, M. E., F. L. Liljestrand, W. C. Hockaday, and C. A. Masiello, Plant species, not climate, controls aboveground biomass $O_2 : CO_2$ exchange ratios in deciduous and coniferous ecosystems, *Journal of Geophysical Research: Biogeosciences*, *122*, 2017JG003,847, 2017.

Gerbig, C., J. C. Lin, J. W. Munger, and S. C. Wofsy, What can tracer observations in the continental boundary layer tell us about surface-atmosphere fluxes?, *Atmospheric Chemistry and Physics*, *6*, 539–554, 2006.

Harvard University, Physical and biological characteristics of the Harvard Forest, 2011, `http://harvardforest.fas.harvard.edu/research/HF-tract` [Accessed 05-01-2018].

Jacobs, A. F. G., J. H. van Boxel, and R. H. Shaw, The dependence of canopy layer turbulence on within-canopy thermal stratification, *Agricultural and Forest Meteorology*, *58*, 247–256, 1992.

Keeling, R., and A. Manning, Studies of recent changes in atmospheric O_2 content, in *Treatise on Geochemistry*, second ed., pp. 385–404, Elsevier, 2014.

Keeling, R. F., Measuring correlations between atmospheric oxygen and carbon dioxide mole fractions: A preliminary study in urban air, *Journal of Atmospheric Chemistry*, *7*, 153–176, 1988.

Keeling, R. F., S. C. Piper, and M. Heimann, Global and hemispheric CO_2 sinks deduced from changes in atmospheric O_2 concentration, *Nature*, *381*, 218–221, 1996.

Keeling, R. F., T. Blaine, B. Paplawsky, L. Katz, C. Atwood, and T. Brockwell, Measurments of changes in atmospheric Ar/N_2 ratio using a rapid-switching, single-capillary mass spectrometer system, *Tellus*, *56B*, 322–338, 2004.

Lee, X., Air motion within and above forest vegetation in non-ideal conditions, *Forest Ecology and Management*, *135*, 3–18, 2000.

Lin, J. C., C. Gerbig, S. C. Wofsy, A. E. Andrews, B. C. Daube, K. J. Davis, and C. A. Grainger, A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, *Journal of Geophysical Research: Atmospheres*, *108*, 4493, 2003.

Munger, J. W., Harvard Forest EMS flux tower data, 2018, `http://atmos.seas.harvard.edu/lab/data/nigec-data.html` [Accessed 05-01-2018].

NOAA, READY gridded meteorological data archives, 2017, `https://ready.arl.noaa.gov/archives.php` [Accessed 04-27-2018].

NOAA, HYSPLIT Hybrid Single Particle Lagrangian Integrated Trajectory model, 2018, `https://ready.arl.noaa.gov/HYSPLIT.php` [Accessed 04-27-2018].

Potosnak, M. J., S. C. Wofsy, A. S. Denning, T. J. Conway, J. W. Munger, and D. H. Barnes, Influence of biotic exchange and combustion sources on atmospheric CO_2 concentrations in New England from observations at a forest flux tower, *Journal of Geophysical Research*, *104*, 9561–9569, 1999.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *13.8: Spectral Analysis of Unevenly Sampled Data*, pp. 685–692, third ed., Cambridge University Press, Cambridge, 2007.

Raupach, M. R., J. J. Finnigan, and Y. Brunet, Coherent eddies and turbulence in vegetation canopies: the mixing-layer analogy, in *Boundary-Layer Meteorology 25th Anniversary Volume, 19701995*, pp. 351–382, Springer, Dordrecht, 1996.

Seibt, U., W. Brand, M. Heimann, J. Lloyd, J. Severinghaus, and L. Wingate, Observations of $O_2 : CO_2$ exchange ratios during ecosystem gas exchange, *Global Biogeochemical Cycles*, *18*, 1–18, 2004.

Severinghaus, J. P., Studies of the terrestrial O_2 and carbon cycles in sand dune gases and in Biosphere 2, Ph.D. thesis, Columbia University, New York, 1995.

Stein, A. F., R. R. Draxler, G. D. Rolph, B. J. B. Stunder, M. D. Cohen, and F. Ngan, NOAAs HYSPLIT atmospheric transport and dispersion modeling system, *Bulletin of the American Meteorological Society*, *96*, 2059–2077, 2015.

Steinbach, J., C. Gerbig, C. Rdenbeck, U. Karstens, C. Minejima, and H. Mukai, The CO_2 release and Oxygen uptake from Fossil Fuel Emission Estimate (COFFEE) dataset: effects from varying oxidative ratios, *Atmospheric Chemistry and Physics*, *11*, 6855–6870, 2011.

Stephens, B. B., P. S. Bakwin, P. P. Tans, R. M. Teclaw, and D. D. Baumann, Application of a differential fuel-cell analyzer for measuring atmospheric oxygen variations, *Journal of Atmospheric and Oceanic Technology*, *24*, 82–94, 2007.

Stull, R., The atmospheric boundary layer, in *Atmospheric Science: An Introductory Survey*, edited by J. M. Wallace and P. V. Hobbs, second ed., chap. 9, pp. 375–417, Academic Press, New York, 2006.

Velasco, E., and M. Roth, Cities as net sources of CO_2 : Review of atmospheric CO_2 exchange in urban environments measured by eddy covariance technique, *Geography Compass*, *4*, 1238–1259, 2010.

Ward, H. C., S. Kotthaus, C. S. B. Grimmond, A. Bjorkegren, M. Wilkinson, W. T. J. Morrison, J. G. Evans, J. I. L. Morison, and M. Iamarino, Effects of urban density on carbon dioxide exchanges: Observations of dense urban, suburban and woodland areas of southern England, *Environmental Pollution*, *198*, 186–200, 2015.

Worrall, F., G. D. Clay, C. A. Masiello, and G. Mynheer, Estimating the oxidative ratio of the global terrestrial biosphere carbon, *Biogeochemistry*, *115*, 23–32, 2013.